THE
LONDON, EDINBURGH AND DUBLIN
PHILOSOPHICAL MAGAZINE
AND
JOURNAL OF SCIENCE

CONDUCTED BY

SIR LAWRENCE BRAGG, O.B.E., M.A., D.Sc., F.R.S.
PROF. G. P. THOMSON, M.A., D.Sc., F.R.S.
AND
ALLAN FERGUSON, M.A., D.Sc.

> "Nec aranearum sane textus ideo melior quia ex se fila gignunt, nec noster vilior quia ex allenis libamus ut apes."
>
> JUST. LIPS. *Polit.* lib. i. cap. l. Not.

VOL. XXXIII.—SEVENTH SERIES
JANUARY—DECEMBER 1942

LONDON:
TAYLOR & FRANCIS, LTD., RED LION COURT, FLEET STREET, E.C. 4.
Agents: WM. DAWSON & SONS, LTD., 129 CANNON STREET, E.C. 4.

"Meditationis est perscrutari occulta; contemplationis est admirari perspicua . . . . Admiratio generat quæstionem, quæstio investigationem, investigatio inventionem."—*Hugo de S. Victore.*

—— "Cur spirent venti, cur terra dehiscat,
Cur mare turgescat, pelago cur tantus amaror,
Cur caput obscura Phœbus ferrugine condat,
Quid toties diros cogat flagrare cometas,
Quid pariat nubes, veniant cur fulmina cœlo,
Quo micet igne Iris, superos quis conciat orbes
Tam vario motu."

J. B. Pinelli ad Mazonium.

ALERE FLAMMAM.

CONTENTS OF VOL. XXXIII. (Nos. 216–227)

(SEVENTH SERIES)

NUMBER 216—JANUARY

Page

I. Numerical Differentiation near the Limits of a Difference Table. By W. G. Bickley, D.Sc., Imperial College of Science and Technology, and J. C. P. Miller, Ph.D., University of Liverpool 1

II. The Pressure Wave System in the Periodic Flow of Liquid in a Pipe. By R. S. Silver, M.A., B.Sc., Ph.D., Chief Physicist, G. & J. Weir, Ltd., Engineers, Cathcart, Glasgow, S. 4 15

III. Dimensions of Physical Quantities. By Prof. W. Wilson, F.R.S. 26

IV. The Steady Flow of Ground-Water to a Pumped Well in the Vicinity of a River. By N. S. Boulton, M.Sc., Assoc.M.Inst.C.E., University of Sheffield 34

V. The Birth and Descent of John Mayow : A Tercentenary Note. By Douglas McKie, D.Sc., Ph.D., Lecturer in the History of Science, University College, London 51

VI. The Influence of Diffusion on the Propagation of Shock Waves. By T. G. Cowling, M.A., D.Phil. (Manchester) 61

VII. The Sedimentation of Washed Red Blood-Cells. By R. B. Whittington, M.Sc. (of the Engineering Department, The University, Manchester) 68

VIII. Notices respecting New Books :—
  - Dr. Friedrich Ellinger's The Biologic Fundamentals of Radiation Therapy 77
  - Mr. Linus Pauling's The Nature of the Chemical Bond 77
  - Dr. Charles Singer's A Short History of Science to the Nineteenth Century 78
  - Dr. Barrer's Diffusion in and through Solids 78
  - Messrs. Bruce Chalmers and A. G. Quarrell's The Physical Examination of Metals—Vol. II. Electrical Methods 79
  - Dr. O. K. Buros's The Second Yearbook of Research and Statistical Methodology Books and Reviews 80

NUMBER 217—FEBRUARY

IX. The Thermal and Electrical Properties of Bitumastic Compounds containing Quartz Sand. By Prof. Willis Jackson, Electro-Technics Department, Manchester University 81

X. Some Comments on the Article by C. G. Pendse. "LXXXVII. A Note on the Definition and Determination of Mass in Newtonian Mechanics."—Phil. Mag., Nov. (Suppl.) 1937. By Glen H. Draper 90

Page

XI. Observations on Mr. Glen H. Draper's Paper entitled "Some Comments on the Article by C. G. Pendse: LXXXVII. A Note on the Definitions and Determination of Mass in Newtonian Mechanics."—Phil. Mag., Nov. (Suppl.) 1937. By C. G. Pendse 92

XII. Collision Relations. By L. A. Pars 96

XIII. Corresponding Problems in Periodic and Steady Flow. By George Green, D.Sc., Lecturer in Applied Physics in the University of Glasgow 102

XIV. Approximate Formulæ for Functions expressed as Definite Integrals. By A. Majid Mian, Ph.D., and S. Chapman, F.R.S. 115

XV. The Forces on an Electron, according to Dirac's Wave Equation. By A. Lees, M.Sc., Ph.D., University College, Swansea 131

XVI. On Yule's Method of Investigating Periodicities of Disturbed Series. The Motion of a Pendulum in a Turbulent Fluid. By T. E. W. Schumann, Ph.D. 138

XVII. The Comma in Music—II. By A. F. Dufton, M.A., D.I.C. 151

XVIII. The Comma in Music.—III. By A. F. Dufton, M.A., D.I.C. .. 155

XIX. Heat Flow when the Boundary Condition is Newton's Law. By J. H. Awbery, B.A., B.Sc., F.Inst.P. 157

XX. The Fermi-Dirac Functions. By F. C. Auluck, Dyal Singh College, Lahore 159

NUMBER 218—MARCH

XXI. Corresponding Problems in Periodic and Steady Flow. By George Green, D.Sc., Lecturer in Applied Physics in the University of Glasgow 161

XXII. The Transient Behaviour of Four-Terminal Networks. By Louis A. Pipes, Ph.D., Graduate School of Engineering, Harvard University, Cambridge, Massachusetts 174

XXIII. Relativity Theory of Electromagnetism. By Miss J. Cattermole, Ph.D, ... 215

XXIV. The Formula of the Selenium Barrier-Layer Photo-Cell. By R. A. Houstoun, D.Sc., University of Glasgow 226

XXV. Notices respecting New Books:—
  Mr. Ruel V. Churchill's Fourier Series and Boundary Value Problems 238
  Mr. R. P. Bell's Acid-Base Catalysis 238

Errata (Ser. 7, vol. xx. September 1935):—
  Dr. N. S. Japolosky's Paper on a Theory of Elementary Particles... 240

NUMBER 219—APRIL

XXVI. The Stresses in an Incomplete Ring subject to Uniform Shear and Compression on its Inner Surface. By S. T. Newing, B.Sc., A.R.C.S., D.I.C. 241

XXVII. The Variation of Viscosity of Liquids with Temperature By M. K. Srinivasan and B. Prasad 258

Page

XXVIII. Non-Linear Distortion, with Particular Reference to the Theory of Frequency Modulated Waves.—Part II. By E. C. Cherry, M.Sc., and R. S. Rivlin, M.A. 272

XXIX. Note on the Bending of Thin Uniformly Loaded Plates bounded by Cardioids, Lemniscates, and certain other Quartic Curves. By Bibhutibhusan Sen, Islamia College, Bengal 294

XXX. Heat Conduction in a Semi-infinite Slab. By W. A. Mersman. 303

XXXI. Effusion Phenomena in a Degenerate Bose-Einstein Gas. By D. V. Gogate, M.Sc., Ph.D. (London) and Y. V. Kathavate, B.Sc., Physics Department, Baroda College, Baroda (India) 310

XXXII. The Pressure Wave System in the Periodic Flow of Liquid in a Pipe. By R. S. Silver, M.A., B.Sc., Ph.D., Chief Physicist, G. & J. Weir, Ltd., Engineers, Cathcart, Glasgow, S. 4 315

XXXIII. Notices respecting New Books :—

Messrs. Samuel Glasstone, Keith J. Laidler, and Henry Eyring's The Theory of Rate Processss 319

NUMBER 220—MAY

XXXIV. On the Dimensions of Physical Magnitudes. By Herbert Dingle, A.R.C.S., D.I.C., D.Sc. 321

XXXV. An Investigation of the Time-Lag in an Electromagnetic Clutch. By F. Horner, M.Sc., Electrotechnics Department, Manchester University 345

XXXVI. Some Disputed Questions in the Philosophy of the Physical Sciences. By E. T. Whittaker 353

XXXVII. A Note on Prof. W. Wilson's "Dimensions of Physical Quantities." By G. Burniston Brown 367

XXXVIII. The Theory of the Electric Charge and the Quantum Theory.—Part III. By H. T. Flint, Reader in Physics, University of London, King's College 369

XXXIX. Structure and Capacity of Electrical Double Layer. By J. J. Bikerman 384

XL. Dimensions. By Norman Campbell, Sc.D., F.Inst.P. 398

NUMBER 221—JUNE

XLI. Contact Potentials.—Part I. General Principles. By J. Alan Chalmers, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham 399

XLII. Contact Potentials.—Part II. The Volta Effect. By J. Alan Chalmers, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham 416

XLIII. Measurement of a Mutual Inductance and Resistance by a Ballistic Method. By E. Tyler, D.Sc., M.Sc., F.Inst.P., A.M.I.E.E., A.F.R.Ae.S., Head of the Science Department, South-East Essex Technical College, Dagenham 431

Page

XLIV. Asymptotic Solutions of Linear Differential Equations. By HAROLD JEFFREYS, F.R.S. 451

XLV. Differentially Cyclical Sets of Functions. An Extension of the Concept of Hyperbolic Functions. By LUDWIK SILBERSTEIN, Ph.D. 457

XLVI. Eigenfunctions in Heisenberg approximation of the Two-electrons Problem. By L. PINCHERLE, King's College, London, at the H. H. Wills' Laboratory, University of Bristol 462

XLVII. The Drifting Force on a Ship among Waves. By T. H. HAVELOCK, F.R.S. 467

XLVIII. Newtonian Mass, its Definition and Determination. By GLEN H. DRAPER 476

NUMBER 222—JULY

XLIX. Units and Dimensions. By E. A. GUGGENHEIM, M.A., Sc.D. 479

L. Contact Potentials.—Part III. The Voltaic Cell. By J. ALAN CHALMERS, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham 496

LI. Contact Potentials.—Part IV. The Concentration Cell. By J. ALAN CHALMERS, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham 506

LII. The Transverse Vibrations of a Helical Spring with Pinned Ends and no Axial Load. By J. DICK, B.Sc., Ph.D. 513

LIII. The Vibratory Electron in Electrostatics. By E. TAYLOR JONES, D.Sc., Professor of Natural Philosophy in the University of Glasgow 519

LIV. Heat Conduction in a Wedge, or an Infinite Cylinder whose Cross-section is a Circle or a Sector of a Circle. By J. C. JAEGER 527

LV. The Space Relation among Five Points in Elliptic and Hyperbolic Geometry. By LUDWIK SILBERSTEIN, Ph.D. 536

LVI. Comments on T. E. W. Schumann's paper, "An Investigation concerning G. I. Taylor's Correlation Coefficient of Turbulence." By CHAIM L. PEKERIS, D.Sc. 541

LVII. A Dynamical Treatment of the Elements of Heat. By G. BURNISTON BROWN, M.Sc., Ph.D., University College, London 543

LVIII. Experiments with Helmholtz Resonators. By H. MARY BROWNING, M.Sc., Ph.D., F.Inst.P., University College, Nottingham 551

LIX. Notices respecting New Books:—
- Mr. GEORGE GAMOW's The Birth and Death of the Sun 556
- Messrs. R. H. HEALEY and J. W. REED's The Behaviour of Slow Electrons in Gases 557

NUMBER 223—AUGUST

LX. The Development of the Trichromatic Theory of Colour Vision. By Prof. W. PEDDIE 559

LXI. The Design of Pole Pieces of Electromagnets and the Forces acting on Small Bodies placed in their Magnetic Fields. By N. DAVY, University College, Nottingham 575

Page

LXII. Contact Potentials.—Part V. The Metal-Electrolyte Volta Effect. By J. Alan Chalmers, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham 594

LXIII. Contact Potentials.—Part VI. Single Potential Differences. By J. Alan Chalmers, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham 599

LXIV. Contact Potentials.—Part VII. Chemical and Thermal Effects. By J. Alan Chalmers, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham 608

LXV. The Application of the Laplace Transformation to a Problem on Elastic Vibrations. By C. J. Tranter, M.A. 614

LXVI. Ultra-Violet Dispersion of Air. By H. Lowery, D.Sc., F.Inst.P. (Plate I.) 622

LXVII. Dimensions. By Prof. W. Wilson, F.R.S. 631

LXVIII. Symmetric Relations between the Coefficients of Reversed Power Series. By W. E. Bleick, Ph.D., U.S. Naval Academy 637

NUMBER 224—SEPTEMBER

LXIX. On the Equilibrium of Plates. By A. C. Stevenson 639

LXX. The Temperature Distribution around a Spherical Hole in an Infinite Conducting Medium. By H. Ll. D. Pugh, B.Sc., and A. J. Harris, B.Sc. 661

LXXI. The Damping of the Heaving and Pitching Motion of a Ship. By T. H. Havelock, F.R.S. 666

LXXII. Relativistic Aspect of the Stress Tensor of the Electromagnetic Field. By Jessie Cattermole, Ph.D. 674

LXXIII. A General Superposition Theorem of the “Thevenin” Type. By G. F. Freeman, M.Sc.(Eng.), A.M.I.E.E., A.M.I.Mech.E., West Ham Municipal College 679

LXXIV. The Viscosity-Temperature Function of Liquids. By E. P. Irany 685

LXXV. Approximate Expressions for Transmission Line Constants in Terms of the Open- and Closed-circuit Impedances. By J. C. Simmonds .. 688

LXXVI. On the Dimensions of Physical Magnitudes. (Second Paper.) By Herbert Dingle, A.R.C.S., D.I.C., D.Sc. 692

LXXVII. Erratum to the Paper “The Formula of the Selenium Barrier-Layer Photo-cell.” By R. A. Houstoun 699

LXXVIII. Notices respecting New Books:—
  Dr. S. Tolansky's Introduction to Atomic Physics 699
  Mr. Alexander Findlay's Introduction to Physical Chemistry 700
  Sir William Cecil Dampier's A History of Science and its Relations with Philosophy and Religion 701
  Reports on Progress in Physics, Vol. VIII. (1941), edited by Dr. W. B. Mann 702
  Dr. A. G. Gaydon's Spectroscopy and Combustion Theory .. 702

NUMBER 225—OCTOBER

Page

LXXIX. Elastic Waves and Vibrations of Thin Rods. By JOHN PRESCOTT, M.A., D.Sc. 703

LXXX. Some New Values for the Exponential Integral. By C. A. COULSON and W. E. DUNCANSON 754

LXXXI. Dimensions and the Facts of Measurement. By NORMAN CAMPBELL 761

LXXXII. Electric Determination of Film-volume and Film-density. By H. LÖWY, Dr.Phil. 772

NUMBER 226—NOVEMBER

LXXXIII. Some Notes and Suggestions on the Teaching of Physics. By C. J. SMITH, Ph.D., A.R.C.S., Lecturer in Physics, Royal Holloway College (University of London), Englefield Green, Surrey 775

LXXXIV. Probability and Quantum Theory. By HAROLD JEFFREYS. 815

LXXXV. Thermodynamical Notes. By Prof. W. WILSON, F.R.S. 831

LXXXVI. Note on Dimensions. By Prof. W. WILSON, F.R.S. 842

LXXXVII. A Universal Valve Voltmeter with Linear Scale. By W. F. LOVERING, B.Sc., University College, Nottingham 844

NUMBER 227—DECEMBER

LXXXVIII. Some Notes on Newton's Chemical Philosophy written upon the Occasion of the Tercentenary of his Birth. By DOUGLAS MCKIE, D.Sc., Ph.D. 847

LXXXIX. Newton and the 'Principia.' By Prof. ALLAN FERGUSON, M.A., D.Sc. 871

XC. The Force required to give a Small Acceleration to a Slowly-moving Sphere carrying a Surface Charge of Electricity. By G. F. C. SEARLE, Sc.D., F.R.S., Cavendish Laboratory, Cambridge 889

XCI. Note on Einstein's Theory of Gravitation. By R. A. HOUSTOUN, D.Sc., University of Glasgow 899

XCII. The Determination of High-Frequency Transmission Line Constants by Measurements at Special Frequencies. By J. C. SIMMONDS. 904

XCIII. Supplementary Note on Newtonian Mass, its Definition and Determination. By GLEN H. DRAPER 910

PLATE

I. Illustrative of Dr. H. LOWERY's Paper on Ultra-Violet Dispersion of Air.

9774-10329

THE

LONDON, EDINBURGH & DUBLIN

PHILOSOPHICAL MAGAZINE

AND

JOURNAL OF SCIENCE

[SEVENTH SERIES—VOL. 33]

I. *Numerical Differentiation near the Limits of a Difference Table.*

By W. G. Bickley, D.Sc., Imperial College of Science and Technology,
and
J. C. P. Miller, Ph.D., University of Liverpool *.

[Received August 26, 1940.]

1. *Introduction.*

This paper has its immediate origin in the implications of a remark by Bradfield and Southwell (1). From the values of a function calculated (approximately) for eleven equally spaced arguments they wished to calculate derivatives of the function for the same arguments, using the known finite difference formulæ. They remark that "the order of the highest derivative which can be calculated is different at different sections"—implying that this order is that of the highest difference which can be used. If the finite difference formulæ are restricted to those containing only forward, backward, or central differences respectively, the implication is correct—but there must exist formulæ of a "mixed"

---

[*] Communicated by the Authors. Copies of this paper are available from the *publishers*, price 1*s*. each, plus postage.

type which enable the whole ten differences to be utilized at any argument, *i. e.*, formulæ commencing as central difference formulæ, switching over to forward (or backward) differences when the edge of the difference table is reached *. The route through the difference table would then be $\longrightarrow\nearrow$ or $\longrightarrow\searrow$. Similar considerations apply near the beginning or end of any difference table. We are not aware that the necessary coefficients have hitherto been published, and the object of this paper is to supply them.

In section 2 the main results, in the form of tables of coefficients, are described, together with (it is hoped sufficient) discussion and examples to enable those who so wish to use the tables without mastering or reading the underlying theory. In the last section the theory, calculation, and checking of the formulæ are considered.

Numerical differentiation—by *any* formula—is of necessity relatively inaccurate when the values of the function are not exact (*e. g.*, rounded-off or experimental), and the smaller the tabular interval the greater is the relative effect of such inexactitude. It is therefore desirable to use a large tabular interval, even though this entails the use of differences of highish order ; we give coefficients for differences up to the twelfth. This inaccuracy increases as one proceeds to higher derivatives, when it is usually combined with a progressive loss of significant figures ; we have therefore refrained from giving complete tables for derivatives of higher order than the fourth, although they have been computed for all orders up to the twelfth. Should they be needed, they can be constructed by applying the rules developed in section 3.3 to the entries of the basic Tables I. and IV.

The possible and/or probable errors in the derivatives calculated from the "mixed" type formulæ vary with the argument. A discussion of them is desirable, and is in hand.

2. *The Tables and the Method of Use.*

2.1. *Notation.*—The usual † notations for forward, backward, and central differences, as exhibited in the following schemes, will be used ;

---

* Formulæ starting diagonally, and finishing horizontally, will also exist. They are less convenient and more laborious in use, and will not be considered here.

† These differ from astronomical usage (Comrie (2)) ; there forward, backward, and central differences are not distinguished by different symbols, Δ being used for central differences, while mean differences are identifiable by difference of parity between index and doubled suffix.

the actual numbers in the same relative positions will, of course, be identical.

x.	Forward.		Backward.		Centr	
a	y_0		y_0		y_0	
	Δy_0		∇y_1		$\delta y_{\frac{1}{2}}$	
$a+w$	y_1	$\Delta^2 y_0$	y_1	$\nabla^2 y_2$	y_1	$\delta^2 y_1$
	Δy_1		∇y_2		$\delta y_{1\frac{1}{2}}$	
$a+2w$	y_2	$\Delta^2 y_1$	y_2	$\nabla^2 y_3$	y_2	$\delta^2 y_2$
	Δy_2		∇y_3		$\delta y_{2\frac{1}{2}}$	
$a+3w$	y_3	$\Delta^2 y_2$	y_3	$\nabla^2 y_4$	y_3	$\delta^2 y_3$
	Δy_3		∇y_4		$\delta y_{3\frac{1}{2}}$	
$a+4w$	y_4		y_4		y_4	

along with
$$\mu\delta^n y_m = \tfrac{1}{2}(\delta^n y_{m+\frac{1}{2}} + \delta^n y_{m-\frac{1}{2}}).$$

The tabular interval is denoted by w. The differential operator will be denoted by D, and we shall use the abbreviation

$$(\mathrm{D}^q y)_{x=a+mw} = \mathrm{D}^q y_m, \text{ or sometimes } \mathrm{D}^q_m.$$

We abbreviate $\delta^n y_m$ similarly to δ^n_m.

2.2. *Description of the Tables.*—Table I. (p. 11) gives the Gregory–Newton coefficients $\mathrm{N}_{q,n}$ in the formulæ

$$w^q\,\mathrm{D}^q y_m = \begin{cases} \sum\limits_{n=q}^{\infty} (-)^{n-q}\,\mathrm{N}_{q,n}\,\Delta^n y_m \\ \sum\limits_{n=q}^{\infty} \mathrm{N}_{q,n}\,\nabla^n y_m \end{cases}$$

for $q=0(1)12$ and $n \not> 12$.

Table II. (p. 12) gives the corresponding central difference coefficients $\mathrm{C}_{q,n}$ in the formulæ

$$w^q\,\mathrm{D}^q y_m = \begin{cases} \sum\limits_{r=0}^{\infty} (-)^r\,\mathrm{C}_{q,q+2r}\,\delta^{q+2r} y_m & (q \text{ even}) \\ \sum\limits_{r=0}^{\infty} (-)^r\,\mathrm{C}_{q,q+2r}\,\mu\delta^{q+2r} y_m & (q \text{ odd}) \end{cases}$$

Tables III., 1–4 (facing p. 12), give the coefficients in the mixed type formulæ for the first four derivatives at the tabulated values of the argument. Their arrangement is such that the coefficient to be used as a multiplier occupies the same relative position as the difference which it is to multiply. The "route" is indicated by the arrows; central difference coefficients are given in bolder type; the central difference coefficient (always unity) on the left of the table for D^q multiplies $\delta^q y_0$ if q is even and $\mu\delta^q y_0$ if q is odd. In order further to clarify the mode of usage of the tables, illustrative examples are worked out in detail (see pp. 4–6).

Tables IV. (p. 13) and V. (p. 14), for arguments at mid-interval, correspond to Tables I. and II., and give $N'_{q,n}$ and $C'_{q,n}$ for the formulæ

$$w^q D^q y_{m-\frac{1}{2}} = \begin{cases} \sum\limits_{n=q}^{\infty} (-)^{n-q} N'_{q,n} \Delta^n y_m \\ \sum\limits_{n=q}^{\infty} N'_{q,n} \nabla^n y_{m-1} \\ \sum\limits_{r=0}^{\infty} (-)^r C'_{q,q+2r} \mu\delta^{q+2r} y_{m-\frac{1}{2}} & (q \text{ even}) \\ \sum\limits_{r=0}^{\infty} (-)^r C'_{q,q+2r} \delta^{q+2r} y_{m-\frac{1}{2}} & (q \text{ odd}) \end{cases}$$

Tables VI., 0–3 (facing p. 14), give the coefficients in the mixed type formulæ for the function and its first three derivatives at mid-interval.

2.3. *Numerical Examples.*—As material we employ values of the function cos x (x in radians) at interval 0·1, to ten decimals. The values and differences are given on p. 5.

We will use these differences, and the coefficients from Tables III. 1, III. 2, and VI. 1, to calculate y' (0·2), y'' (0·2), and y' (0·25) respectively.

(i) y' (0·2)

Coefficient (Table III., 1)	×	difference (p. 5)	=	term in result
+1		$\mu\delta_2 =$		−1983 38381
−1/6		$\mu\delta^3_2 =$ 19 81732		− 3 30288 7
+1/30		$\Delta^5_0 = \delta^5_{2\frac{1}{2}} = -$ 24686		− 822 9
−1/60		$\Delta^6_0 = \delta^6_3 = -$ 9533		+ 158 9
+1/105		$\Delta^7_0 = \delta^7_{3\frac{1}{2}} = +$ 347		+ 3 3
−1/168		$\Delta^8_0 = \delta^8_4 = +$ 81		− 5
				−1986 69330 9

Since $w=0·1$, this means that

$$y'\ (0·2) = -0·19866\ 9331.$$

The exact value is −sin 0·2=−0·19866 93307 95 . . .

(ii) y'' (0·2)

Coefficient (Table III., 2)	×	difference (p. 5)	=	term in result
+ 1		$\delta^2_2 =$		−979 25012
− 1/12		$\delta^4_2 = +$9 78432		− 81536 0
+ 1/90		$\Delta^6_0 = \delta^6_3 = -$ 9533		− 105 9
− 1/90		$\Delta^7_0 = \delta^7_{3\frac{1}{2}} = +$ 347		− 3 9
+47/5040		$\Delta^8_0 = \delta^8_4 = +$ 81		+ 8
−19/2520		$\Delta^9_0 = \delta^9_{4\frac{1}{2}} = +$ 25		− 2
				−980 06657 2

x	$\cos x$	δ	δ^2	δ^3	δ^4	δ^5	δ^6	δ^7	δ^8	δ^9
0·0	1·00000 00000									
		−499 58347								
0·1	0·99500 41653		−994 17528							
		−1493 75875		+14 92516						
0·2	·98006 65778	**−1983 38381**	−979 25012	**+19 81732**	+9 78432					
		−2473 00887		+24 70948		−24686				
0·3	·95533 64891		−954 54064		+9 53746		−9533			
		−3427 54951		+34 24694		−34219		+347		
0·4	·92106 09940		−920 29370		+9 19527		−9186		+ 81	
		−4347 84321		+43 44221		−43405		+428		+25
0·5	·87758 25619		−876 85149		+8 76122		−8758		+106	
		−5224 69470		+52 20343		−52163		+534		−59
0·6	·82533 56149		−824 64806		+8 23959		−8224		+ 47	
		−6049 34276		+60 44302		−60387		+581		
0·7	·76484 21873		−764 20504		+7 63572		−7643			
		−6813 54780		+68 07874		−68030				
0·8	·69670 67093		−696 12630		+6 95542					
		−7509 67410		+75 03416						
0·9	·62160 99683		−621 09214							
		−8130 76624								
1·0	·54030 23059									

The terms in bold type are mean central differences.

Since $w^2=0{\cdot}01$, this means that

$$y''\,(0{\cdot}2)=-0{\cdot}98006\;657.$$

This may be compared with the true value,

$$-\cos 0{\cdot}2=-0{\cdot}98006\;65778\ldots$$

(iii) $y'\,(0{\cdot}25)$

Coefficient (Table VI., 1)	×	difference (p. 5)	=	term in result
$+1$		$\delta_{2\frac{1}{2}}=$		$-2473\;00887$
$-1/24$		$\delta^3_{2\frac{1}{2}}=+24\;70948$		$-\quad 1\;02956\;2$
$+3/640$		$\delta^5_{2\frac{1}{2}}=-\quad 24686$		$-\quad 115\;7$
$-5/7168$		$\Delta^7_0=\delta^7_{3\frac{1}{2}}=+\quad 347$		$-\quad 2$
$+5/7168$		$\Delta^8_0=\delta^8_4=+\quad 81$		$+\quad 1$
				$-2474\;03959\;0$

Thus $$y'\,(0{\cdot}25)=-0{\cdot}24740\;3959,$$

while $$-\sin 0{\cdot}25=-0{\cdot}24740\;39592\;5\ldots$$

Scrutiny of (i) and (ii) will show how and why decimals, and hence significant figures, are progressively lost as we reduce the interval or as we calculate higher derivatives, whenever the tabular interval is less than unity.

3. *Derivation of Formulæ and Calculation of Coefficients.*

3.1. *Gregory–Newton Coefficients.*—Denoting as usual by E the operation of adding w to the argument, or its equivalent of increasing any suffix by unity,

$$y_1=\mathrm{E}y_0=(1+\Delta)y_0=e^{w\mathrm{D}}y_0$$

where the last expression is the symbolic form of Taylor's series. Thus

$$\mathrm{E}=(1+\Delta)=e^{w\mathrm{D}}$$

so that $$w\mathrm{D}=\log\,(1+\Delta)$$

$$=\Delta-\tfrac{1}{2}\Delta^2+\tfrac{1}{3}\Delta^3-\tfrac{1}{4}\Delta^4+\ldots$$

Consequently

$$w^q\mathrm{D}^q y_m=(\Delta-\tfrac{1}{2}\Delta^2+\tfrac{1}{3}\Delta^3-\tfrac{1}{4}\Delta^4+\ldots)^q y_m$$

$$=\sum_{n=q}^{\infty}(-)^{n-q}\mathrm{N}_{q,\,n}\Delta^n y_m$$

Similarly $$\mathrm{E}^{-1}=(1-\nabla)=e^{-w\mathrm{D}}$$

and $$w\mathrm{D}=-\log\,(1-\nabla)$$

$$=\nabla+\tfrac{1}{2}\nabla^2+\tfrac{1}{3}\nabla^3+\tfrac{1}{4}\nabla^4+\ldots$$

so that
$$w^q D^q y_m = (\nabla + \tfrac{1}{2}\nabla^2 + \tfrac{1}{3}\nabla^3 + \tfrac{1}{4}\nabla^4 + \dots)^q y_m$$
$$= \sum_{n=q}^{\infty} N_{q,n} \nabla^n y_m$$

The coefficients $N_{q,n}$ can thus be calculated by involution of the logarithmic series. Other—and less laborious—methods exist. They depend on the Gregory–Newton interpolation formula

$$y(a+rw) = y_r = \Sigma \frac{r(r-1)(r-2)\dots(r-n+1)}{n!} \Delta^n y_0$$

Differentiating q times with respect to r gives

$$w^q D^q y_r = \Sigma \frac{d^q\{r(r-1)(r-2)\dots(r-n+1)\}/dr^q}{n!} \Delta^n y_0$$

so that

$$n!\, N_{q,n} = (-)^{n-q}[d^q\{r(r-1)(r-2)\dots(r-n+1)\}/dr^q]_{r=0}$$
$$= [d^q\{r(r+1)(r+2)\dots(r+n-1)\}/dr^q]_{r=0}$$

which is essentially Markoff's formula. The values of the derivatives when $r=0$ can evidently be found as simple multiples of the coefficients, and this leads to a second method of calculating $N_{q,n}$.

From the last result we can also derive a third method. Thus

$$(n+1)!\, N_{q,n+1} = [(d/dr)^q\{r(r+1)(r+2)\dots(r+n)\}]_{r=0}$$
$$= [(r+n)(d/dr)^q\{r(r+1)\dots(r+n-1)\}$$
$$+ q(d/dr)^{q-1}\{r(r+1)(r+2)\dots(r+n-1)\}]_{r=0}$$

or, on reduction,

$$(n+1)N_{q,n+1} = nN_{q,n} + qN_{q-1,n}$$

This recurrence formula, coupled with the fact that $N_{q,q}=1$, enables Table I. to be built up column by column, and provides what is probably the quickest and least laborious way of doing so. Actually all three methods have been employed, and they check one another.

It has seemed worth while also to provide formulæ for mid-interval, since in numerical processes the time often comes when a reduction of the interval is imperative, and the simplest method of so doing is to halve it. The corresponding coefficients are obtained from formulæ very similar to those above.

Putting $r = s - \frac{1}{2}$, the Gregory–Newton formula becomes

$$y_{s-\frac{1}{2}} = \sum_{n=0}^{\infty} \frac{(s-\frac{1}{2})(s-\frac{3}{2})\dots(s-n+\frac{1}{2})}{n!} \Delta^n y_0$$

whence we derive, as in the case of $N_{q,n}$

$$n!\, N'_{q,n} = [(d/ds)^q\{(s+\tfrac{1}{2})(s+\tfrac{3}{2})\dots(s+n-\tfrac{1}{2})\}]_{s=0}$$

and
$$2(n+1)N'_{q,n+1} = (2n+1)N'_{q,n} + 2qN'_{q-1,n}$$

This recurrence formula, coupled with $N'_{q,q}=1$, $N'_{q,0}=0$ $(q>0)$, enables Table IV. to be constructed. From the above it is also evident that the N' are multiples of the coefficients in the polynomial

$$(x+1)(x+3)\ldots(x+2n-1),$$

which provides a second method of calculating them. A third method depends on the fact that

$$w^qD^qy_{-\frac{1}{2}}=(\Delta-\tfrac{1}{2}\Delta^2+\tfrac{1}{3}\Delta^3-\ldots)^q(1+\Delta)^{-\frac{1}{2}}y_0$$

3.2. *Central Difference Formulæ.*—These have been given, up to the terms involving the twelfth difference, by Comrie (2). It will transpire in our next subsection that the coefficients automatically emerge in the process of calculating the coefficients for the "mixed" formulæ. To calculate them independently we have

$$\delta=E^{\frac{1}{2}}-E^{-\frac{1}{2}}=2\sinh\tfrac{1}{2}wD$$

so that

$$wD=2\sinh^{-1}(\delta/2)$$

Also

$$\mu=\tfrac{1}{2}(E^{\frac{1}{2}}+E^{-\frac{1}{2}})=\cosh\tfrac{1}{2}wD=\sqrt{(1+\tfrac{1}{4}\delta^2)}$$

Consequently

$$w^qD^q=\begin{cases}\{2\sinh^{-1}(\delta/2)\}^q & \begin{cases}q\text{ even, interval points}\\ q\text{ odd, mid-interval}\end{cases}\\[2ex] \mu\{2\sinh^{-1}(\delta/2)\}^q(1+\tfrac{1}{4}\delta^2)^{-\frac{1}{2}} & \begin{cases}q\text{ odd, interval points}\\ q\text{ even, mid-interval}\end{cases}\end{cases}$$

For further information and details concerning differentiation formulæ see Steffensen (4) §7, and §18.

3.3. *Mixed Formulæ.*—Although the general rule by which the coefficients in the "mixed" formulæ are derived is the same in all cases, its deduction will probably be more easily followed if we analyse the particular case of the first derivative.

We have

$$\begin{aligned}wDy_1&=(\Delta-\tfrac{1}{2}\Delta^2+\tfrac{1}{3}\Delta^3-\tfrac{1}{4}\Delta^4\ldots)y_1\\&=\Delta y_1-(\tfrac{1}{2}\Delta^2-\tfrac{1}{3}\Delta^3+\tfrac{1}{4}\Delta^4\ldots)(1+\Delta)y_0\\&=\Delta y_1-\tfrac{1}{2}\Delta^2y_0+\{(\tfrac{1}{3}-\tfrac{1}{2})\Delta^3-(\tfrac{1}{4}-\tfrac{1}{3})\Delta^4\ldots\}y_0\\&=\mu\delta y_1-\tfrac{1}{6}\Delta^3y_0+\tfrac{1}{12}\Delta^4y_0-\ldots\end{aligned}$$

since

$$\mu\delta y_1=\tfrac{1}{2}(\Delta y_1+\Delta y_0)=\Delta y_1-\tfrac{1}{2}\Delta^2y_0$$

Similarly

$$\begin{aligned}wDy_2&=\mu\delta y_2-(\tfrac{1}{6}\Delta^3-\tfrac{1}{12}\Delta^4+\tfrac{1}{20}\Delta^5\ldots)y_1\\&=\mu\delta y_2-\tfrac{1}{6}\Delta^3y_1+(\tfrac{1}{12}\Delta^4-\tfrac{1}{20}\Delta^5\ldots)(1+\Delta)y_0\\&=\mu\delta y_2-\tfrac{1}{6}\mu\delta^3y_1+\{(\tfrac{1}{12}-\tfrac{1}{20})\Delta^5-(\tfrac{1}{20}-\tfrac{1}{30})\Delta^6+\ldots\}y_0\end{aligned}$$

$$\begin{array}{ccccc}
1 & & d_2 & & \\
& c_1 & & d_3 & \\
1 & & c_2 & & d_4 \\
& b_1 & & c_3 & \\
1 & & b_2 & & c_4 \\
& a_1 & & b_3 & \\
& & a_2 & & b_4 \\
& & & a_3 & \\
& & & & a_4
\end{array}$$

The process can be indefinitely continued, for the law of formation of the coefficients is now clear. If the accompanying table represents the coefficients, with

$$a_n = (-)^n \mathrm{N}_{q,\,n+q}$$

then

$$b_n = a_{n-1} + a_n$$

$$c_n = b_{n-1} + b_n$$

and so on. The upward progress in any column ceases when the corresponding backward Gregory–Newton coefficient is reached. The reproduction of the known central difference coefficients, and later of the backward Gregory–Newton coefficients, provides a check (practically conclusive and complete) upon the accuracy of the calculations.

Modifications are still necessary to convert a table so constructed into the corresponding member of the set III. or VI. In the case of odd derivatives at tabular points, or of even derivatives at mid-interval—that is, where the central difference formula involves mean differences—the central line of coefficients in the original table is repeated three times, at interval $\frac{1}{2}w$, with $\delta_{-\frac{1}{2}}$, $\mu\delta_0$ and $\delta_{\frac{1}{2}}$ respectively in III., 1, **3**, or with δ_0, $\mu\delta_{\frac{1}{2}}$ and δ_1 in VI., 0, 2. In the case of even derivatives at tabular points, or of odd derivatives at mid-interval, the original table has a central line of zeros, and the entries in the lines above and below this are equal. In Tables III. 2, 4, these entries are repeated three times, with δ_{-1}, δ_0, and δ_1 respectively (or with $\delta_{-\frac{1}{2}}$, $\delta_{\frac{1}{2}}$ and $\delta_{1\frac{1}{2}}$ in Tables VI. 1, 3), instead of twice, at interval w; the zeros are not printed in the tables.

The suffixes also need consideration. In Tables III. the sum of the outermost pair in any column is zero. In Tables VI., however, an asymmetry appears, owing to the fact that the tables relate to suffix $+\frac{1}{2}$; the sum of the outermost suffixes in any column is, in this case, unity.

It may also be mentioned that the number of 1's in the left-hand column *as computed* is equal to the order of the derivative, for the interval point schemes, but one more than this for mid-interval schemes.

To illustrate these points we reproduce a few columns of the tables for D^2y and D^3y as computed; these should be compared with the corresponding portions of Tables III. 2, and III. 3.

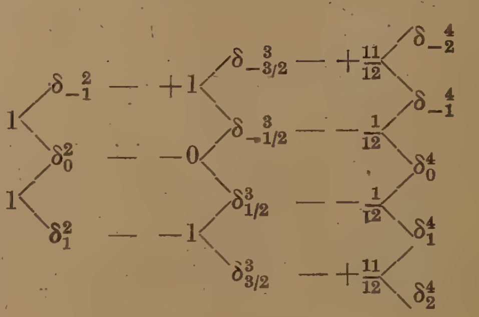
The tables could also have been exhibited in the form of "hexagon" diagrams (Fraser (3)), and portions corresponding to parts of the tables just mentioned are given, to exhibit this possibility. In the hexagon diagrams horizontal lines indicate addition or subtraction, sloping lines indicate multiplication, and any continuous route from left to right may be followed.

D^2y_0

Original Table.

				137/180
			5/6	
		11/12		−13/180
	1		−1/12	
1		−1/12		1/90
	0		0	
1		−1/12		1/90
	−1		1/12	
		11/12		−13/180
			−5/6	
				137/180

Hexagon Diagram.

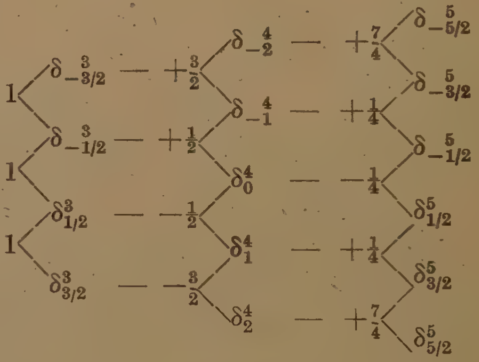

1
δ^2_{-1}
$+1$
$\delta^3_{-3/2}$
$+\frac{11}{12}$
δ^4_{-2}
δ^4_{-1}
δ^2_0
-0
$\delta^3_{-1/2}$
$-\frac{1}{12}$
δ^4_0
1
$\delta^3_{1/2}$
$-\frac{1}{12}$
δ^4_1
δ^2_1
-1
$\delta^3_{3/2}$
$+\frac{11}{12}$
δ^4_2

D^3y_0

Original Table.

				29/15
			15/8	
		7/4		7/120
	3/2		1/8	
1		1/4		−1/15
	1/2		−1/8	
1		−1/4		7/120
	−1/2		1/8	
1		1/4		−1/15
	−3/2		−1/8	
		7/4		7/120
			−15/8	
				29/15

Hexagon Diagram.


$\delta^5_{-5/2}$
δ^4_{-2}
$+\frac{7}{4}$
1
$\delta^3_{-3/2}$
$+\frac{3}{2}$
$\delta^5_{-3/2}$
δ^4_{-1}
$+\frac{1}{4}$
$\delta^3_{-1/2}$
$+\frac{1}{2}$
$\delta^5_{-1/2}$
1
δ^4_0
$-\frac{1}{4}$
$\delta^3_{1/2}$
$-\frac{1}{2}$
$\delta^5_{1/2}$
1
δ^4_1
$+\frac{1}{4}$
$\delta^3_{3/2}$
$-\frac{3}{2}$
$\delta^5_{3/2}$
δ^4_2
$+\frac{7}{4}$
$\delta^5_{5/2}$

Acknowlegements.

The authors wish to record, firstly their thanks to the Royal Society for a grant in aid of the publication of this paper, and secondly their appreciation of the care and courtesy with which the printers have endeavoured to meet their wishes, especially as regards the printing and arrangement of the tables.

References.

(1) Bradfield, K. N. and Southwell, R. V., "Relaxation Methods applied to Engineering Problems.—I." Proc. Roy. Soc. A, clxi. pp. 155–181 (1937).
(2) Comrie, L. J., 'Interpolation and Allied Tables,' London, H.M. Stationery Office (1936). (Reprinted from the Nautical Almanac for 1937.)
(3) Fraser, D. C., "On the Graphic Delineation of Interpolation Formulæ," Journ. Inst. Act. xliii. pp. 235–241 (1909).
(4) Steffensen, J. F., 'Interpolation,' Baltimore (1927).

TABLE I. $N_{q,n}$

q	n	0	1	2	3	4	5	6	7	8	9	10	11	12	q
0		1	0	0	0	0	0	0	0	0	0	0	0	0	0
1			1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{7}$	$\frac{1}{8}$	$\frac{1}{9}$	$\frac{1}{10}$	$\frac{1}{11}$	$\frac{1}{12}$	1
2				1	1	$\frac{11}{12}$	$\frac{5}{6}$	$\frac{137}{180}$	$\frac{7}{10}$	$\frac{363}{560}$	$\frac{761}{1260}$	$\frac{7129}{12600}$	$\frac{671}{1260}$	$\frac{83711}{1\,66320}$	2
3					1	$\frac{3}{2}$	$\frac{7}{4}$	$\frac{15}{8}$	$\frac{29}{15}$	$\frac{469}{240}$	$\frac{29531}{15120}$	$\frac{1303}{672}$	$\frac{16103}{8400}$	$\frac{1\,90553}{1\,00800}$	3
4						1	2	$\frac{17}{6}$	$\frac{7}{2}$	$\frac{967}{240}$	$\frac{89}{20}$	$\frac{4523}{945}$	$\frac{7645}{1512}$	$\frac{3\,41747}{64800}$	4
5							1	$\frac{5}{2}$	$\frac{25}{6}$	$\frac{35}{6}$	$\frac{1069}{144}$	$\frac{285}{32}$	$\frac{31063}{3024}$	$\frac{1\,39381}{12096}$	5
6								1	3	$\frac{23}{4}$	9	$\frac{3013}{240}$	$\frac{781}{48}$	$\frac{2\,42537}{12096}$	6
7									1	$\frac{7}{2}$	$\frac{91}{12}$	$\frac{105}{8}$	$\frac{4781}{240}$	$\frac{13321}{480}$	7
8										1	4	$\frac{29}{3}$	$\frac{55}{3}$	$\frac{10831}{360}$	8
9											1	$\frac{9}{2}$	12	$\frac{99}{4}$	9
10												1	5	$\frac{175}{12}$	10
11													1	$\frac{11}{2}$	11
12														1	12

$$w^q D^q y_m = \sum_{n=q}^{\infty} (-)^{n-q} N_{q,n} \Delta^n y_m = \sum_{n=q}^{\infty} N_{q,n} \nabla^n y_m$$

$$N_{q,q}=1 \quad N_{0,n}=0 \quad (n>0) \quad N_{1,n}=1/n \quad (n+1)N_{q,n+1}=nN_{q,n}+qN_{q-1,n}$$

TABLE II. $C_{q,n}$

q \ n	0	1	2	3	4	5	6	7	8	9	10	11	12	q
0	1		0		0		0		0		0		0	0
1		1		$\frac{1}{6}$		$\frac{1}{30}$		$\frac{1}{140}$		$\frac{1}{630}$		$\frac{1}{2772}$		1
2			1		$\frac{1}{12}$		$\frac{1}{90}$		$\frac{1}{560}$		$\frac{1}{3150}$		$\frac{1}{16632}$	2
3				1		$\frac{1}{4}$		$\frac{7}{120}$		$\frac{41}{3024}$		$\frac{479}{1\,51200}$		3
4					1		$\frac{1}{6}$		$\frac{7}{240}$		$\frac{41}{7560}$		$\frac{479}{4\,53600}$	4
5						1		$\frac{1}{3}$		$\frac{13}{144}$		$\frac{139}{6048}$		5
6							1		$\frac{1}{4}$		$\frac{13}{240}$		$\frac{139}{12096}$	6
7								1		$\frac{5}{12}$		$\frac{31}{240}$		7
8									1		$\frac{1}{3}$		$\frac{31}{360}$	8
9										1		$\frac{1}{2}$		9
10											1		$\frac{5}{12}$	10
11												1		11
12													1	12

$$w^q D^q y_m = \begin{cases} \sum_{r=0}^{\infty} (-)^r C_{q,q+2r}\, \delta^{q+2r} y_m & (q \text{ even}) \\ \sum_{r=0}^{\infty} (-)^r C_{q,q+2r}\, \mu\delta^{q+2r} y_m & (q \text{ odd}) \end{cases}$$

TABLE III. 1. Dy_0

																	$+\frac{1}{12}\delta^{12}_{-6}$
																$+\frac{1}{11}\delta^{11}_{-5\frac{1}{2}}$	
														$+\frac{1}{10}\delta^{10}_{-5}$			$-\frac{1}{132}\delta^{12}_{-5}$
													$+\frac{1}{9}\delta^{9}_{-4\frac{1}{2}}$			$-\frac{1}{110}\delta^{11}_{-4\frac{1}{2}}$	
											$+\frac{1}{8}\delta^{8}_{-4}$			$-\frac{1}{90}\delta^{10}_{-4}$			$+\frac{1}{660}\delta^{12}_{-4}$
										$+\frac{1}{7}\delta^{7}_{-3\frac{1}{2}}$			$-\frac{1}{72}\delta^{9}_{-3\frac{1}{2}}$			$+\frac{1}{495}\delta^{11}_{-3\frac{1}{2}}$	
								$+\frac{1}{6}\delta^{6}_{-3}$			$-\frac{1}{56}\delta^{8}_{-3}$			$+\frac{1}{360}\delta^{10}_{-3}$			$-\frac{1}{1980}\delta^{12}_{-3}$
							$+\frac{1}{5}\delta^{5}_{-2\frac{1}{2}}$			$-\frac{1}{42}\delta^{7}_{-2\frac{1}{2}}$			$+\frac{1}{252}\delta^{9}_{-2\frac{1}{2}}$			$-\frac{1}{1320}\delta^{11}_{-2\frac{1}{2}}$	
					$+\frac{1}{4}\delta^{4}_{-2}$			$-\frac{1}{30}\delta^{6}_{-2}$			$+\frac{1}{168}\delta^{8}_{-2}$			$-\frac{1}{840}\delta^{10}_{-2}$			$+\frac{1}{3960}\delta^{12}_{-2}$
				$+\frac{1}{3}\delta^{3}_{-1\frac{1}{2}}$			$-\frac{1}{20}\delta^{5}_{-1\frac{1}{2}}$			$+\frac{1}{105}\delta^{7}_{-1\frac{1}{2}}$			$-\frac{1}{504}\delta^{9}_{-1\frac{1}{2}}$			$+\frac{1}{2310}\delta^{11}_{-1\frac{1}{2}}$	
		$+\frac{1}{2}\delta^{2}_{-1}$			$-\frac{1}{12}\delta^{4}_{-1}$			$+\frac{1}{60}\delta^{6}_{-1}$			$-\frac{1}{280}\delta^{8}_{-1}$			$+\frac{1}{1260}\delta^{10}_{-1}$			$-\frac{1}{5544}\delta^{12}_{-1}$
	$+1\ \delta_{-\frac{1}{2}}$			$-\frac{1}{6}\delta^{3}_{-\frac{1}{2}}$			$+\frac{1}{30}\delta^{5}_{-\frac{1}{2}}$			$-\frac{1}{140}\delta^{7}_{-\frac{1}{2}}$			$+\frac{1}{630}\delta^{9}_{-\frac{1}{2}}$			$-\frac{1}{2772}\delta^{11}_{-\frac{1}{2}}$	
$\rightarrow$	$+1\mu\delta_0$		$\rightarrow$	$-\frac{1}{6}\mu\delta^{3}_{0}$		$\rightarrow$	$+\frac{1}{30}\mu\delta^{5}_{0}$		$\rightarrow$	$-\frac{1}{140}\mu\delta^{7}_{0}$		$\rightarrow$	$+\frac{1}{630}\mu\delta^{9}_{0}$		$\rightarrow$	$-\frac{1}{2772}\mu\delta^{11}_{0}$	
	$+1\ \delta_{\frac{1}{2}}$			$-\frac{1}{6}\delta^{3}_{\frac{1}{2}}$			$+\frac{1}{30}\delta^{5}_{\frac{1}{2}}$			$-\frac{1}{140}\delta^{7}_{\frac{1}{2}}$			$+\frac{1}{630}\delta^{9}_{\frac{1}{2}}$			$-\frac{1}{2772}\delta^{11}_{\frac{1}{2}}$	
		$-\frac{1}{2}\delta^{2}_{1}$			$+\frac{1}{12}\delta^{4}_{1}$			$-\frac{1}{60}\delta^{6}_{1}$			$+\frac{1}{280}\delta^{8}_{1}$			$-\frac{1}{1260}\delta^{10}_{1}$			$+\frac{1}{5544}\delta^{12}_{1}$
				$+\frac{1}{3}\delta^{3}_{1\frac{1}{2}}$			$-\frac{1}{20}\delta^{5}_{1\frac{1}{2}}$			$+\frac{1}{105}\delta^{7}_{1\frac{1}{2}}$			$-\frac{1}{504}\delta^{9}_{1\frac{1}{2}}$			$+\frac{1}{2310}\delta^{11}_{1\frac{1}{2}}$	
					$-\frac{1}{4}\delta^{4}_{2}$			$+\frac{1}{30}\delta^{6}_{2}$			$-\frac{1}{168}\delta^{8}_{2}$			$+\frac{1}{840}\delta^{10}_{2}$			$-\frac{1}{3960}\delta^{12}_{2}$
							$+\frac{1}{5}\delta^{5}_{2\frac{1}{2}}$			$-\frac{1}{42}\delta^{7}_{2\frac{1}{2}}$			$+\frac{1}{252}\delta^{9}_{2\frac{1}{2}}$			$-\frac{1}{1320}\delta^{11}_{2\frac{1}{2}}$	
								$-\frac{1}{6}\delta^{6}_{3}$			$+\frac{1}{56}\delta^{8}_{3}$			$-\frac{1}{360}\delta^{10}_{3}$			$+\frac{1}{1980}\delta^{12}_{3}$
										$+\frac{1}{7}\delta^{7}_{3\frac{1}{2}}$			$-\frac{1}{72}\delta^{9}_{3\frac{1}{2}}$			$+\frac{1}{495}\delta^{11}_{3\frac{1}{2}}$	
											$-\frac{1}{8}\delta^{8}_{4}$			$+\frac{1}{90}\delta^{10}_{4}$			$-\frac{1}{660}\delta^{12}_{4}$
													$+\frac{1}{9}\delta^{9}_{4\frac{1}{2}}$			$-\frac{1}{110}\delta^{11}_{4\frac{1}{2}}$	
														$-\frac{1}{10}\delta^{10}_{5}$			$+\frac{1}{132}\delta^{12}_{5}$
																$+\frac{1}{11}\delta^{11}_{5\frac{1}{2}}$	
																	$-\frac{1}{12}\delta^{12}_{6}$

TABLE III. 2. D^2y_0

										$+\frac{83711}{1\,66320}\delta^{12}_{-6}$
									$+\frac{671}{1260}\delta^{11}_{-5\frac{1}{2}}$	
								$+\frac{7129}{12600}\delta^{10}_{-5}$		$-\frac{4861}{1\,66320}\delta^{12}_{-5}$
							$+\frac{761}{1260}\delta^{9}_{-4\frac{1}{2}}$		$-\frac{419}{12600}\delta^{11}_{-4\frac{1}{2}}$	
						$+\frac{363}{560}\delta^{8}_{-4}$		$-\frac{481}{12600}\delta^{10}_{-4}$		$+\frac{3349}{8\,31600}\delta^{12}_{-4}$
					$+\frac{7}{10}\delta^{7}_{-3\frac{1}{2}}$		$\frac{223}{5040}\delta^{9}_{-3\frac{1}{2}}$		$+\frac{31}{6300}\delta^{11}_{-3\frac{1}{2}}$	
				$+\frac{137}{180}\delta^{6}_{-3}$		$-\frac{29}{560}\delta^{8}_{-3}$		$+\frac{17}{2800}\delta^{10}_{-3}$		$-\frac{743}{8\,31600}\delta^{12}_{-3}$
			$+\frac{5}{6}\delta^{5}_{-2\frac{1}{2}}$		$-\frac{11}{180}\delta^{7}_{-2\frac{1}{2}}$		$+\frac{19}{2520}\delta^{9}_{-2\frac{1}{2}}$		$-\frac{29}{25200}\delta^{11}_{-2\frac{1}{2}}$	
		$+\frac{11}{12}\delta^{4}_{-2}$		$-\frac{13}{180}\delta^{6}_{-2}$		$+\frac{47}{5040}\delta^{8}_{-2}$		$-\frac{37}{25200}\delta^{10}_{-2}$		$+\frac{107}{4\,15800}\delta^{12}_{-2}$
	$+1\,\delta^{3}_{-1\frac{1}{2}}$		$-\frac{1}{12}\delta^{5}_{-1\frac{1}{2}}$		$+\frac{1}{90}\delta^{7}_{-1\frac{1}{2}}$		$-\frac{1}{560}\delta^{9}_{-1\frac{1}{2}}$		$+\frac{1}{3150}\delta^{11}_{-1\frac{1}{2}}$	
$+1\,\delta^{2}_{-1}$		$-\frac{1}{12}\delta^{4}_{-1}$		$+\frac{1}{90}\delta^{6}_{-1}$		$-\frac{1}{560}\delta^{8}_{-1}$		$+\frac{1}{3150}\delta^{10}_{-1}$		$-\frac{1}{16632}\delta^{12}_{-1}$
$+1\,\delta^{2}_{0}$		$-\frac{\mathbf{1}}{\mathbf{12}}\delta^{4}_{0}$		$+\frac{\mathbf{1}}{\mathbf{90}}\delta^{6}_{0}$		$-\frac{\mathbf{1}}{\mathbf{560}}\delta^{8}_{0}$		$+\frac{\mathbf{1}}{\mathbf{3150}}\delta^{10}_{0}$		$-\frac{\mathbf{1}}{\mathbf{16632}}\delta^{12}_{0}$
$+1\,\delta^{2}_{1}$		$-\frac{1}{12}\delta^{4}_{1}$		$+\frac{1}{90}\delta^{6}_{1}$		$-\frac{1}{560}\delta^{8}_{1}$		$+\frac{1}{3150}\delta^{10}_{1}$		$-\frac{1}{16632}\delta^{12}_{1}$
	$-1\,\delta^{3}_{1\frac{1}{2}}$		$+\frac{1}{12}\delta^{5}_{1\frac{1}{2}}$		$-\frac{1}{90}\delta^{7}_{1\frac{1}{2}}$		$+\frac{1}{560}\delta^{9}_{1\frac{1}{2}}$		$-\frac{1}{3150}\delta^{11}_{1\frac{1}{2}}$	
		$+\frac{11}{12}\delta^{4}_{2}$		$-\frac{13}{180}\delta^{6}_{2}$		$+\frac{47}{5040}\delta^{8}_{2}$		$-\frac{37}{25200}\delta^{10}_{2}$		$+\frac{107}{4\,15800}\delta^{12}_{2}$
			$-\frac{5}{6}\delta^{5}_{2\frac{1}{2}}$		$+\frac{11}{180}\delta^{7}_{2\frac{1}{2}}$		$-\frac{19}{2520}\delta^{9}_{2\frac{1}{2}}$		$+\frac{29}{25200}\delta^{11}_{2\frac{1}{2}}$	
				$+\frac{137}{180}\delta^{6}_{3}$		$-\frac{29}{560}\delta^{8}_{3}$		$+\frac{17}{2800}\delta^{10}_{3}$		$-\frac{743}{8\,31600}\delta^{12}_{3}$
					$-\frac{7}{10}\delta^{7}_{3\frac{1}{2}}$		$+\frac{223}{5040}\delta^{9}_{3\frac{1}{2}}$		$-\frac{31}{6300}\delta^{11}_{3\frac{1}{2}}$	
						$+\frac{363}{560}\delta^{8}_{4}$		$-\frac{481}{12600}\delta^{10}_{4}$		$+\frac{3349}{8\,31600}\delta^{12}_{4}$
							$-\frac{761}{1260}\delta^{9}_{4\frac{1}{2}}$		$+\frac{419}{12600}\delta^{11}_{4\frac{1}{2}}$	
								$+\frac{7129}{12600}\delta^{10}_{5}$		$-\frac{4861}{1\,66320}\delta^{12}_{5}$
									$-\frac{671}{1260}\delta^{11}_{5\frac{1}{2}}$	
										$+\frac{83711}{1\,66320}\delta^{12}_{6}$

TABLE III. 3. D^3y_0

									$+\frac{1\,90553}{1\,00800}\delta^{12}_{-6}$
								$+\frac{16103}{8400}\delta^{11}_{-5\frac12}$	
							$+\frac{1303}{672}\delta^{10}_{-5}$		$-\frac{2683}{1\,00800}\delta^{12}_{-5}$
						$+\frac{29531}{15120}\delta^{9}_{-4\frac12}$		$-\frac{123}{5600}\delta^{11}_{-4\frac12}$	
					$+\frac{469}{240}\delta^{8}_{-4}$		$-\frac{61}{4320}\delta^{10}_{-4}$		$-\frac{97}{14400}\delta^{12}_{-4}$
				$+\frac{29}{15}\delta^{7}_{-3\frac12}$		$-\frac{1}{945}\delta^{9}_{-3\frac12}$		$-\frac{593}{75600}\delta^{11}_{-3\frac12}$	
			$+\frac{15}{8}\delta^{6}_{-3}$		$+\frac{1}{48}\delta^{8}_{-3}$		$-\frac{79}{6048}\delta^{10}_{-3}$		$+\frac{193}{60480}\delta^{12}_{-3}$
		$+\frac{7}{4}\delta^{5}_{-2\frac12}$		$+\frac{7}{120}\delta^{7}_{-2\frac12}$		$\frac{331}{15120}\delta^{9}_{-2\frac12}$		$\frac{263}{50400}\delta^{11}_{-2\frac12}$	
	$+\frac{3}{2}\delta^{4}_{-2}$		$+\frac{1}{8}\delta^{6}_{-2}$		$-\frac{3}{80}\delta^{8}_{-2}$		$+\frac{89}{10080}\delta^{10}_{-2}$		$\frac{613}{3\,02400}\delta^{12}_{-2}$
$+1\ \delta^{3}_{-1\frac12}$		$-\frac{1}{4}\delta^{5}_{-1\frac12}$		$-\frac{1}{15}\delta^{7}_{-1\frac12}$		$+\frac{59}{3780}\delta^{9}_{-1\frac12}$		$\frac{13}{3000}\delta^{11}_{-1\frac12}$	
	$+\frac{1}{2}\delta^{4}_{-1}$		$-\frac{1}{8}\delta^{6}_{-1}$		$+\frac{7}{240}\delta^{8}_{-1}$		$-\frac{41}{6048}\delta^{10}_{-1}$		$\frac{479}{3\,02400}\delta^{12}_{-1}$
$+1\ \delta^{3}_{-\frac12}$		$\frac{1}{4}\delta^{5}_{-\frac12}$		$+\frac{7}{120}\delta^{7}_{-\frac12}$		$-\frac{41}{3024}\delta^{9}_{-\frac12}$		$\frac{479}{1\,51200}\delta^{11}_{-\frac12}$	
$+1\mu\delta^{3}_{0}$		$-\frac{1}{4}\mu\delta^{5}_{0}$		$+\frac{7}{120}\mu\delta^{7}_{0}$		$-\frac{41}{3024}\mu\delta^{9}_{0}$		$\frac{479}{1\,51200}\mu\delta^{11}_{0}$	
$+1\ \delta^{3}_{\frac12}$		$-\frac{1}{4}\delta^{5}_{\frac12}$		$+\frac{7}{120}\delta^{7}_{\frac12}$		$-\frac{41}{3024}\delta^{9}_{\frac12}$		$+\frac{479}{1\,51200}\delta^{11}_{\frac12}$	
	$-\frac{1}{2}\delta^{4}_{1}$		$+\frac{1}{8}\delta^{6}_{1}$		$-\frac{7}{240}\delta^{8}_{1}$		$+\frac{41}{6048}\delta^{10}_{1}$		$\frac{479}{3\,02400}\delta^{12}_{1}$
$+1\ \delta^{3}_{1\frac12}$		$+\frac{1}{4}\delta^{5}_{1\frac12}$		$\frac{1}{15}\delta^{7}_{1\frac12}$		$+\frac{59}{3780}\delta^{9}_{1\frac12}$		$-\frac{13}{3000}\delta^{11}_{1\frac12}$	
	$-\frac{3}{2}\delta^{4}_{2}$		$-\frac{1}{8}\delta^{6}_{2}$		$+\frac{3}{80}\delta^{8}_{2}$		$-\frac{89}{10080}\delta^{10}_{2}$		$\frac{613}{3\,02400}\delta^{12}_{2}$
		$+\frac{7}{4}\delta^{5}_{2\frac12}$		$+\frac{7}{120}\delta^{7}_{2\frac12}$		$-\frac{331}{15120}\delta^{9}_{2\frac12}$		$-\frac{263}{50400}\delta^{11}_{2\frac12}$	
			$-\frac{15}{8}\delta^{6}_{3}$		$-\frac{1}{48}\delta^{8}_{3}$		$+\frac{79}{6048}\delta^{10}_{3}$		$-\frac{193}{60480}\delta^{12}_{3}$
				$+\frac{29}{15}\delta^{7}_{3}$		$-\frac{1}{945}\delta^{9}_{3\frac12}$		$-\frac{593}{75600}\delta^{11}_{3\frac12}$	
					$-\frac{469}{240}\delta^{8}_{4}$		$+\frac{61}{4320}\delta^{10}_{4}$		$+\frac{97}{14400}\delta^{12}_{4}$
						$+\frac{29531}{15120}\delta^{9}_{4\frac12}$		$-\frac{123}{5600}\delta^{11}_{4\frac12}$	
							$-\frac{1303}{672}\delta^{10}_{5}$		$+\frac{2683}{1\,00800}\delta^{12}_{5}$
								$+\frac{16103}{8400}\delta^{11}_{5\frac12}$	
									$-\frac{1\,90553}{1\,00800}\delta^{12}_{6}$

TABLE III. 4. D^4y_0

								$+\frac{3\,41747}{64800}\delta^{12}_{-6}$
							$+\frac{7645}{1512}\delta^{11}_{-5\frac12}$	
						$+\frac{4523}{945}\delta^{10}_{-5}$		$+\frac{98729}{4\,53600}\delta^{12}_{-5}$
					$+\frac{89}{20}\delta^{9}_{-4\frac12}$		$+\frac{2041}{7560}\delta^{11}_{-4\frac12}$	
				$+\frac{967}{240}\delta^{8}_{-4}$		$+\frac{1271}{3780}\delta^{10}_{-4}$		$-\frac{23731}{4\,53600}\delta^{12}_{-4}$
			$+\frac{7}{2}\delta^{7}_{-3\frac12}$		$+\frac{101}{240}\delta^{9}_{-3\frac12}$		$-\frac{167}{2520}\delta^{11}_{-3\frac12}$	
		$+\frac{17}{6}\delta^{6}_{-3}$		$+\frac{127}{240}\delta^{8}_{-3}$		$-\frac{1279}{15120}\delta^{10}_{-3}$		$+\frac{6329}{4\,53600}\delta^{12}_{-3}$
	$+2\,\delta^{5}_{-2\frac12}$		$+\frac{2}{3}\delta^{7}_{-2\frac12}$		$-\frac{13}{120}\delta^{9}_{-2\frac12}$		$+\frac{277}{15120}\delta^{11}_{-2\frac12}$	
$+1\,\delta^{4}_{-2}$		$+\frac{5}{6}\delta^{6}_{-2}$		$-\frac{11}{80}\delta^{8}_{-2}$		$+\frac{359}{15120}\delta^{10}_{-2}$		$-\frac{283}{64800}\delta^{12}_{-2}$
	$+1\,\delta^{5}_{-1\frac12}$		$-\frac{1}{6}\delta^{7}_{-1\frac12}$		$+\frac{7}{240}\delta^{9}_{-1\frac12}$		$-\frac{41}{7560}\delta^{11}_{-1\frac12}$	
$+1\,\delta^{4}_{-1}$		$-\frac{1}{6}\delta^{6}_{-1}$		$+\frac{7}{240}\delta^{8}_{-1}$		$-\frac{41}{7560}\delta^{10}_{-1}$		$+\frac{479}{4\,53600}\delta^{12}_{-1}$
$+1\,\delta^{4}_{0}$		$-\frac{1}{6}\delta^{6}_{0}$		$+\frac{7}{240}\delta^{8}_{0}$		$-\frac{41}{7560}\delta^{10}_{0}$		$+\frac{479}{4\,53600}\delta^{12}_{0}$
$+1\,\delta^{4}_{1}$		$-\frac{1}{6}\delta^{6}_{1}$		$+\frac{7}{240}\delta^{8}_{1}$		$-\frac{41}{7560}\delta^{10}_{1}$		$+\frac{479}{4\,53600}\delta^{12}_{1}$
	$-1\,\delta^{5}_{1\frac12}$		$+\frac{1}{6}\delta^{7}_{1\frac12}$		$-\frac{7}{240}\delta^{9}_{1\frac12}$		$+\frac{41}{7560}\delta^{11}_{1\frac12}$	
$+1\,\delta^{4}_{2}$		$+\frac{5}{6}\delta^{6}_{2}$		$-\frac{11}{80}\delta^{8}_{2}$		$+\frac{359}{15120}\delta^{10}_{2}$		$-\frac{283}{64800}\delta^{12}_{2}$
	$-2\,\delta^{5}_{2\frac12}$		$-\frac{2}{3}\delta^{7}_{2\frac12}$		$+\frac{13}{120}\delta^{9}_{2\frac12}$		$-\frac{277}{15120}\delta^{11}_{2\frac12}$	
		$+\frac{17}{6}\delta^{6}_{3}$		$+\frac{127}{240}\delta^{8}_{3}$		$-\frac{1279}{15120}\delta^{10}_{3}$		$+\frac{6329}{4\,53600}\delta^{12}_{3}$
			$-\frac{7}{2}\delta^{7}_{3\frac12}$		$-\frac{101}{240}\delta^{9}_{3\frac12}$		$+\frac{167}{2520}\delta^{11}_{3\frac12}$	
				$+\frac{967}{240}\delta^{8}_{4}$		$+\frac{1271}{3780}\delta^{10}_{4}$		$-\frac{23731}{4\,53600}\delta^{12}_{4}$
					$-\frac{89}{20}\delta^{9}_{4\frac12}$		$-\frac{2041}{7560}\delta^{11}_{4\frac12}$	
						$+\frac{4523}{945}\delta^{10}_{5}$		$+\frac{98729}{4\,53600}\delta^{12}_{5}$
							$-\frac{7645}{1512}\delta^{11}_{5\frac12}$	
								$+\frac{3\,41747}{64800}\delta^{12}_{6}$

TABLE IV. $N'_{q,n}$

q \ n	0	1	2	3	4	5	6	7	8	9	10	11	12	q
0	1	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{5}{16}$	$\frac{35}{128}$	$\frac{63}{256}$	$\frac{231}{1024}$	$\frac{429}{2048}$	$\frac{6435}{32768}$	$\frac{12155}{65536}$	$\frac{46189}{2\,62144}$	$\frac{88179}{5\,24288}$	$\frac{6\,76039}{41\,94304}$	0
1		1	1	$\frac{23}{24}$	$\frac{11}{12}$	$\frac{563}{640}$	$\frac{1627}{1920}$	$\frac{88069}{1\,07520}$	$\frac{1423}{1792}$	$\frac{15\,93269}{20\,64384}$	$\frac{77\,59469}{103\,21920}$	$\frac{317\,30711}{432\,53760}$	$\frac{465\,22243}{648\,80640}$	1
2			1	$\frac{3}{2}$	$\frac{43}{24}$	$\frac{95}{48}$	$\frac{12139}{5760}$	$\frac{25333}{11520}$	$\frac{81227}{35840}$	$\frac{4\,98233}{2\,15040}$	$\frac{1215\,63469}{516\,09600}$	$\frac{2461\,83839}{1032\,19200}$	$\frac{3\,28081\,17961}{1\,36249\,34400}$	2
3				1	2	$\frac{23}{8}$	$\frac{29}{8}$	$\frac{8197}{1920}$	$\frac{2317}{480}$	$\frac{51\,42611}{9\,67680}$	$\frac{11\,11619}{1\,93536}$	$\frac{3161\,11237}{516\,09600}$	$\frac{5005\,69373}{774\,14400}$	3
4					1	$\frac{5}{2}$	$\frac{101}{24}$	$\frac{287}{48}$	$\frac{14861}{1920}$	$\frac{12103}{1280}$	$\frac{107\,49419}{9\,67680}$	$\frac{245\,63869}{19\,35360}$	$\frac{65980\,23581}{4644\,86400}$	4
5						1	3	$\frac{139}{24}$	$\frac{55}{6}$	$\frac{14927}{1152}$	$\frac{19627}{1152}$	$\frac{41\,24677}{1\,93536}$	$\frac{8\,29385}{32256}$	5
6							1	$\frac{7}{2}$	$\frac{61}{8}$	$\frac{213}{16}$	$\frac{39209}{1920}$	$\frac{1\,10539}{3840}$	$\frac{370\,06861}{9\,67680}$	6
7								1	4	$\frac{233}{24}$	$\frac{445}{24}$	$\frac{58933}{1920}$	$\frac{44359}{960}$	7
8									1	$\frac{9}{2}$	$\frac{289}{24}$	$\frac{1199}{48}$	$\frac{2\,55751}{5760}$	8
9										1	5	$\frac{117}{8}$	$\frac{131}{4}$	9
10											1	$\frac{11}{2}$	$\frac{419}{24}$	10
11												1	6	11
12													1	12

$$\therefore\ w^q\, D^q y_{m-\frac{1}{2}} = \sum_{n=q}^{\infty} (-)^{n-q} N'_{q,n} \Delta^n y_m = \sum_{n=q}^{\infty} N'_{q,n} \nabla^n y_{m-1}$$

$$N'_{q,q}=1 \qquad N'_{0,n}=1.3\ldots(2n-1)/2.4\ldots2n \qquad 2(n+1)N'_{q,n+1}=(2n+1)N'_{q,n}+2qN'_{q-1,n}$$

TABLE V. $C'_{q,n}$

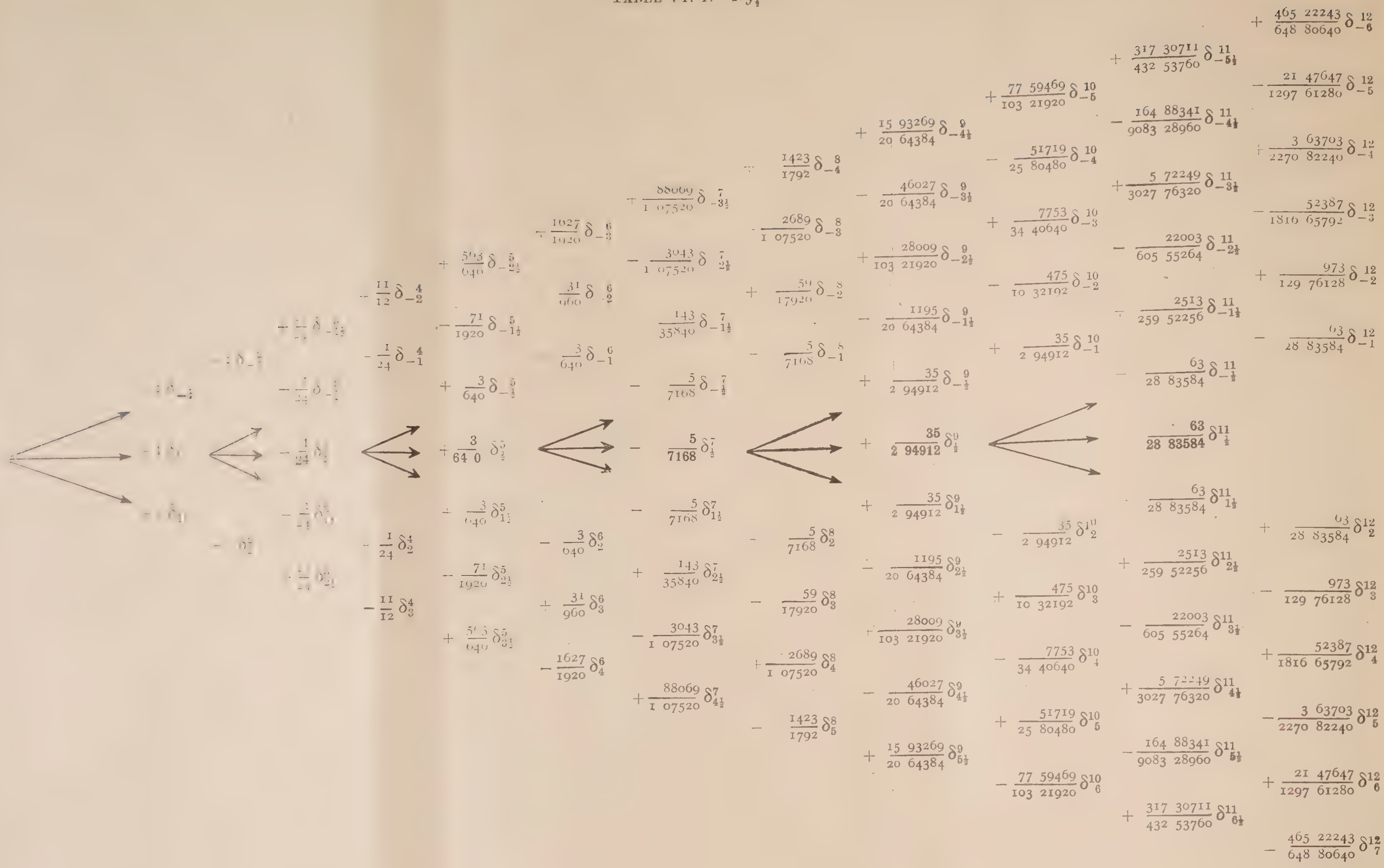
q	n	0	1	2	3	4	5	6	7	8	9	10	11	12	q
0		1		$\frac{1}{8}$		$\frac{3}{128}$		$\frac{5}{1024}$		$\frac{35}{32768}$		$\frac{63}{2\ 62144}$		$\frac{231}{41\ 94304}$	0
1			1		$\frac{1}{24}$		$\frac{3}{640}$		$\frac{5}{7168}$		$\frac{35}{2\ 94912}$		$\frac{63}{28\ 83584}$		1
2				1		$\frac{5}{24}$		$\frac{259}{5760}$		$\frac{3229}{3\ 22560}$		$\frac{1\ 17469}{516\ 09600}$		$\frac{71\ 56487}{1\ 36249\ 34400}$	2
3					1		$\frac{1}{8}$		$\frac{37}{1920}$		$\frac{3229}{9\ 67680}$		$\frac{10679}{172\ 03200}$		3
4						1		$\frac{7}{24}$		$\frac{47}{640}$		$\frac{17281}{9\ 67680}$		$\frac{19\ 97021}{4644\ 86400}$	4
5							1		$\frac{5}{24}$		$\frac{47}{1152}$		$\frac{1571}{1\ 93536}$		5
6								1		$\frac{3}{8}$		$\frac{209}{1920}$		$\frac{28067}{9\ 67680}$	6
7									1		$\frac{7}{24}$		$\frac{133}{1920}$		7
8										1		$\frac{11}{24}$		$\frac{871}{5760}$	8
9											1		$\frac{3}{8}$		9
10												1		$\frac{13}{24}$	10
11													1		11
12														1	12

$$w^q D^q y_{m-\frac{1}{2}} = \begin{cases} \sum\limits_{r=0}^{\infty} (-)^r C'_{q,q+2r} \mu\delta^{q+2r} y_{m-\frac{1}{2}} & (q \text{ even}) \\ \sum\limits_{r=0}^{\infty} (-)^r C'_{q,q+2r} \delta^{q+2r} y_{m-\frac{1}{2}} & (q \text{ odd}) \end{cases}$$

TABLE VI. 0. $y_{\frac{1}{2}}$

												$+\frac{6\,76039}{41\,94304}\delta^{12}_{-6}$
											$+\frac{88179}{5\,24288}\delta^{11}_{-5\frac{1}{2}}$	
										$+\frac{46189}{2\,62144}\delta^{10}_{-5}$		$-\frac{29393}{41\,94304}\delta^{12}_{-5}$
									$+\frac{12155}{65536}\delta^{9}_{-4\frac{1}{2}}$		$-\frac{4199}{5\,24288}\delta^{11}_{-4\frac{1}{2}}$	
								$+\frac{6435}{32768}\delta^{8}_{-4}$		$-\frac{2431}{2\,62144}\delta^{10}_{-4}$		$+\frac{4199}{41\,94304}\delta^{12}_{-4}$
							$+\frac{429}{2048}\delta^{7}_{-3\frac{1}{2}}$		$-\frac{715}{65536}\delta^{9}_{-3\frac{1}{2}}$		$+\frac{663}{5\,24288}\delta^{11}_{-3\frac{1}{2}}$	
						$+\frac{231}{1024}\delta^{6}_{-3}$		$-\frac{429}{32768}\delta^{8}_{-3}$		$+\frac{429}{2\,62144}\delta^{10}_{-3}$		$-\frac{1105}{41\,94304}\delta^{12}_{-3}$
					$+\frac{63}{256}\delta^{5}_{-2\frac{1}{2}}$		$-\frac{33}{2048}\delta^{7}_{-2\frac{1}{2}}$		$+\frac{143}{65536}\delta^{9}_{-2\frac{1}{2}}$		$-\frac{195}{5\,24288}\delta^{11}_{-2\frac{1}{2}}$	
				$+\frac{35}{128}\delta^{4}_{-2}$		$-\frac{21}{1024}\delta^{6}_{-2}$		$+\frac{99}{32768}\delta^{8}_{-2}$		$-\frac{143}{2\,62144}\delta^{10}_{-2}$		$+\frac{455}{41\,94304}\delta^{12}_{-2}$
			$+\frac{5}{16}\delta^{3}_{-1\frac{1}{2}}$		$-\frac{7}{256}\delta^{5}_{-1\frac{1}{2}}$		$+\frac{9}{2048}\delta^{7}_{-1\frac{1}{2}}$		$-\frac{55}{65536}\delta^{9}_{-1\frac{1}{2}}$		$+\frac{91}{5\,24288}\delta^{11}_{-1\frac{1}{2}}$	
		$+\frac{3}{8}\delta^{2}_{-1}$		$-\frac{5}{128}\delta^{4}_{-1}$		$+\frac{7}{1024}\delta^{6}_{-1}$		$-\frac{45}{32768}\delta^{8}_{-1}$		$+\frac{77}{2\,62144}\delta^{10}_{-1}$		$-\frac{273}{41\,94304}\delta^{12}_{-1}$
	$+\frac{1}{2}\delta_{-\frac{1}{2}}$		$-\frac{1}{16}\delta^{3}_{-\frac{1}{2}}$		$+\frac{3}{256}\delta^{5}_{-\frac{1}{2}}$		$-\frac{5}{2048}\delta^{7}_{-\frac{1}{2}}$		$+\frac{35}{65536}\delta^{9}_{-\frac{1}{2}}$		$-\frac{63}{5\,24288}\delta^{11}_{-\frac{1}{2}}$	
$+1\;y_0$		$-\frac{1}{8}\delta^{2}_{0}$		$+\frac{3}{128}\delta^{4}_{0}$		$-\frac{5}{1024}\delta^{6}_{0}$		$+\frac{35}{32768}\delta^{8}_{0}$		$-\frac{63}{2\,62144}\delta^{10}_{0}$		$+\frac{231}{41\,94304}\delta^{12}_{0}$
$+1\,\mu y_{\frac{1}{2}}$		$-\frac{1}{8}\mu\delta^{2}_{\frac{1}{2}}$		$+\frac{3}{128}\mu\delta^{4}_{\frac{1}{2}}$		$-\frac{5}{1024}\mu\delta^{6}_{\frac{1}{2}}$		$+\frac{35}{32768}\mu\delta^{8}_{\frac{1}{2}}$		$-\frac{63}{2\,62144}\mu\delta^{10}_{\frac{1}{2}}$		$+\frac{231}{41\,94304}\mu\delta^{12}_{\frac{1}{2}}$
$+1\;y_1$		$-\frac{1}{8}\delta^{2}_{1}$		$+\frac{3}{128}\delta^{4}_{1}$		$-\frac{5}{1024}\delta^{6}_{1}$		$+\frac{35}{32768}\delta^{8}_{1}$		$-\frac{63}{2\,62144}\delta^{10}_{1}$		$+\frac{231}{41\,94304}\delta^{12}_{1}$
	$-\frac{1}{2}\delta_{1\frac{1}{2}}$		$+\frac{1}{16}\delta^{3}_{1\frac{1}{2}}$		$-\frac{3}{256}\delta^{5}_{1\frac{1}{2}}$		$+\frac{5}{2048}\delta^{7}_{1\frac{1}{2}}$		$-\frac{35}{65536}\delta^{9}_{1\frac{1}{2}}$		$+\frac{63}{5\,24288}\delta^{11}_{1\frac{1}{2}}$	
		$+\frac{3}{8}\delta^{2}_{2}$		$-\frac{5}{128}\delta^{4}_{2}$		$+\frac{7}{1024}\delta^{6}_{2}$		$-\frac{45}{32768}\delta^{8}_{2}$		$+\frac{77}{2\,62144}\delta^{10}_{2}$		$-\frac{273}{41\,94304}\delta^{12}_{2}$
			$-\frac{5}{16}\delta^{3}_{2\frac{1}{2}}$		$+\frac{7}{256}\delta^{5}_{2\frac{1}{2}}$		$-\frac{9}{2048}\delta^{7}_{2\frac{1}{2}}$		$+\frac{55}{65536}\delta^{9}_{2\frac{1}{2}}$		$-\frac{91}{5\,24288}\delta^{11}_{2\frac{1}{2}}$	
				$+\frac{35}{128}\delta^{4}_{3}$		$-\frac{21}{1024}\delta^{6}_{3}$		$+\frac{99}{32768}\delta^{8}_{3}$		$-\frac{143}{2\,62144}\delta^{10}_{3}$		$+\frac{455}{41\,94304}\delta^{12}_{3}$
					$-\frac{63}{256}\delta^{5}_{3\frac{1}{2}}$		$+\frac{33}{2048}\delta^{7}_{3\frac{1}{2}}$		$-\frac{143}{65536}\delta^{9}_{3\frac{1}{2}}$		$+\frac{195}{5\,24288}\delta^{11}_{3\frac{1}{2}}$	
						$+\frac{231}{1024}\delta^{6}_{4}$		$-\frac{429}{32768}\delta^{8}_{4}$		$+\frac{429}{2\,62144}\delta^{10}_{4}$		$-\frac{1105}{41\,94304}\delta^{12}_{4}$
							$-\frac{429}{2048}\delta^{7}_{4\frac{1}{2}}$		$+\frac{715}{65536}\delta^{9}_{4\frac{1}{2}}$		$-\frac{663}{5\,24288}\delta^{11}_{4\frac{1}{2}}$	
								$+\frac{6435}{32768}\delta^{8}_{5}$		$-\frac{2431}{2\,62144}\delta^{10}_{5}$		$+\frac{4199}{41\,94304}\delta^{12}_{5}$
									$-\frac{12155}{65536}\delta^{9}_{5\frac{1}{2}}$		$+\frac{4199}{5\,24288}\delta^{11}_{5\frac{1}{2}}$	
										$+\frac{46189}{2\,62144}\delta^{10}_{6}$		$-\frac{29393}{41\,94304}\delta^{12}_{6}$
											$-\frac{88179}{5\,24288}\delta^{11}_{6\frac{1}{2}}$	
												$+\frac{6\,76039}{41\,94304}\delta^{12}_{7}$

TABLE VI. 1. $Dy_{\frac{1}{2}}$



$+ 1\,\delta_{-\frac{1}{2}}$
$+ 1\,\delta_{\frac{1}{2}}$
$+ 1\,\delta_{1\frac{1}{2}}$

$+ 1\,\delta^{2}_{-1}$
$- 1\,\delta^{2}_{2}$

$+ \frac{23}{24}\,\delta^{3}_{-1\frac{1}{2}}$
$- \frac{1}{24}\,\delta^{3}_{-\frac{1}{2}}$
$- \frac{1}{24}\,\delta^{3}_{\frac{1}{2}}$
$- \frac{1}{24}\,\delta^{3}_{1\frac{1}{2}}$
$+ \frac{23}{24}\,\delta^{3}_{2\frac{1}{2}}$

$+ \frac{11}{12}\,\delta^{4}_{-2}$
$- \frac{1}{24}\,\delta^{4}_{-1}$
$- \frac{1}{24}\,\delta^{4}_{2}$
$- \frac{11}{12}\,\delta^{4}_{3}$

$+ \frac{563}{640}\,\delta^{5}_{-2\frac{1}{2}}$
$- \frac{71}{1920}\,\delta^{5}_{-1\frac{1}{2}}$
$+ \frac{3}{640}\,\delta^{5}_{-\frac{1}{2}}$
$+ \frac{3}{640}\,\delta^{5}_{\frac{1}{2}}$
$+ \frac{3}{640}\,\delta^{5}_{1\frac{1}{2}}$
$- \frac{71}{1920}\,\delta^{5}_{2\frac{1}{2}}$
$+ \frac{563}{640}\,\delta^{5}_{3\frac{1}{2}}$

$+ \frac{1627}{1920}\,\delta^{6}_{-3}$
$\frac{31}{960}\,\delta^{6}_{-2}$
$\frac{3}{640}\,\delta^{6}_{-1}$
$- \frac{3}{640}\,\delta^{6}_{2}$
$+ \frac{31}{960}\,\delta^{6}_{3}$
$- \frac{1627}{1920}\,\delta^{6}_{4}$

$+ \frac{88069}{1\ 07520}\,\delta^{7}_{-3\frac{1}{2}}$
$- \frac{3043}{1\ 07520}\,\delta^{7}_{-2\frac{1}{2}}$
$\frac{143}{35840}\,\delta^{7}_{-1\frac{1}{2}}$
$- \frac{5}{7168}\,\delta^{7}_{-\frac{1}{2}}$
$- \frac{5}{7168}\,\delta^{7}_{\frac{1}{2}}$
$- \frac{5}{7168}\,\delta^{7}_{1\frac{1}{2}}$
$+ \frac{143}{35840}\,\delta^{7}_{2\frac{1}{2}}$
$- \frac{3043}{1\ 07520}\,\delta^{7}_{3\frac{1}{2}}$
$+ \frac{88069}{1\ 07520}\,\delta^{7}_{4\frac{1}{2}}$

$+ \frac{1423}{1792}\,\delta^{8}_{-4}$
$- \frac{2689}{1\ 07520}\,\delta^{8}_{-3}$
$+ \frac{59}{17920}\,\delta^{8}_{-2}$
$- \frac{5}{7168}\,\delta^{8}_{-1}$
$- \frac{5}{7168}\,\delta^{8}_{2}$
$- \frac{59}{17920}\,\delta^{8}_{3}$
$+ \frac{2689}{1\ 07520}\,\delta^{8}_{4}$
$- \frac{1423}{1792}\,\delta^{8}_{5}$

$+ \frac{15\ 93269}{20\ 64384}\,\delta^{9}_{-4\frac{1}{2}}$
$- \frac{46027}{20\ 64384}\,\delta^{9}_{-3\frac{1}{2}}$
$+ \frac{28009}{103\ 21920}\,\delta^{9}_{-2\frac{1}{2}}$
$- \frac{1195}{20\ 64384}\,\delta^{9}_{-1\frac{1}{2}}$
$+ \frac{35}{2\ 94912}\,\delta^{9}_{-\frac{1}{2}}$
$+ \frac{35}{2\ 94912}\,\delta^{9}_{\frac{1}{2}}$
$+ \frac{35}{2\ 94912}\,\delta^{9}_{1\frac{1}{2}}$
$- \frac{1195}{20\ 64384}\,\delta^{9}_{2\frac{1}{2}}$
$+ \frac{28009}{103\ 21920}\,\delta^{9}_{3\frac{1}{2}}$
$- \frac{46027}{20\ 64384}\,\delta^{9}_{4\frac{1}{2}}$
$+ \frac{15\ 93269}{20\ 64384}\,\delta^{9}_{5\frac{1}{2}}$

$+ \frac{77\ 59469}{103\ 21920}\,\delta^{10}_{-5}$
$- \frac{51719}{25\ 80480}\,\delta^{10}_{-4}$
$+ \frac{7753}{34\ 40640}\,\delta^{10}_{-3}$
$- \frac{475}{10\ 32192}\,\delta^{10}_{-2}$
$+ \frac{35}{2\ 94912}\,\delta^{10}_{-1}$
$- \frac{35}{2\ 94912}\,\delta^{10}_{2}$
$+ \frac{475}{10\ 32192}\,\delta^{10}_{3}$
$- \frac{7753}{34\ 40640}\,\delta^{10}_{4}$
$+ \frac{51719}{25\ 80480}\,\delta^{10}_{5}$
$- \frac{77\ 59469}{103\ 21920}\,\delta^{10}_{6}$

$+ \frac{317\ 30711}{432\ 53760}\,\delta^{11}_{-5\frac{1}{2}}$
$- \frac{164\ 88341}{9083\ 28960}\,\delta^{11}_{-4\frac{1}{2}}$
$+ \frac{5\ 72249}{3027\ 76320}\,\delta^{11}_{-3\frac{1}{2}}$
$- \frac{22003}{605\ 55264}\,\delta^{11}_{-2\frac{1}{2}}$
$+ \frac{2513}{259\ 52256}\,\delta^{11}_{-1\frac{1}{2}}$
$- \frac{63}{28\ 83584}\,\delta^{11}_{-\frac{1}{2}}$
$\frac{63}{28\ 83584}\,\delta^{11}_{\frac{1}{2}}$
$\frac{63}{28\ 83584}\,\delta^{11}_{1\frac{1}{2}}$
$+ \frac{2513}{259\ 52256}\,\delta^{11}_{2\frac{1}{2}}$
$- \frac{22003}{605\ 55264}\,\delta^{11}_{3\frac{1}{2}}$
$+ \frac{5\ 72249}{3027\ 76320}\,\delta^{11}_{4\frac{1}{2}}$
$- \frac{164\ 88341}{9083\ 28960}\,\delta^{11}_{5\frac{1}{2}}$
$+ \frac{317\ 30711}{432\ 53760}\,\delta^{11}_{6\frac{1}{2}}$

$+ \frac{465\ 22243}{648\ 80640}\,\delta^{12}_{-6}$
$- \frac{21\ 47647}{1297\ 61280}\,\delta^{12}_{-5}$
$+ \frac{3\ 63703}{2270\ 82240}\,\delta^{12}_{-4}$
$- \frac{52387}{1816\ 65792}\,\delta^{12}_{-3}$
$+ \frac{973}{129\ 76128}\,\delta^{12}_{-2}$
$- \frac{63}{28\ 83584}\,\delta^{12}_{-1}$
$+ \frac{63}{28\ 83584}\,\delta^{12}_{2}$
$- \frac{973}{129\ 76128}\,\delta^{12}_{3}$
$+ \frac{52387}{1816\ 65792}\,\delta^{12}_{4}$
$- \frac{3\ 63703}{2270\ 82240}\,\delta^{12}_{5}$
$+ \frac{21\ 47647}{1297\ 61280}\,\delta^{12}_{6}$
$- \frac{465\ 22243}{648\ 80640}\,\delta^{12}_{7}$

TABLE VI. 2. $D^2y_{\frac12}$

$+1\,\delta^2_{-1}$
$+1\,\delta^2_{0}$
$+1\mu\delta^2_{\frac12}$
$+1\,\delta^2_{1}$
$+1\,\delta^2_{2}$

$+\frac{3}{2}\delta^3_{-1\frac12}$
$+\frac{1}{2}\delta^3_{-\frac12}$
$-\frac{1}{2}\delta^3_{1\frac12}$
$-\frac{3}{2}\delta^3_{2\frac12}$

$+\frac{43}{24}\delta^4_{-2}$
$+\frac{7}{24}\delta^4_{-1}$
$-\frac{5}{24}\delta^4_{0}$
$-\frac{5}{24}\mu\delta^4_{\frac12}$
$-\frac{5}{24}\delta^4_{1}$
$+\frac{7}{24}\delta^4_{2}$
$+\frac{43}{24}\delta^4_{3}$

$+\frac{95}{48}\delta^5_{-2\frac12}$
$+\frac{3}{16}\delta^5_{-1\frac12}$
$-\frac{5}{48}\delta^5_{-\frac12}$
$+\frac{5}{48}\delta^5_{1\frac12}$
$-\frac{3}{16}\delta^5_{2\frac12}$
$-\frac{95}{48}\delta^5_{3\frac12}$

$+\frac{12139}{5760}\delta^6_{-3}$
$+\frac{739}{5760}\delta^6_{-2}$
$-\frac{341}{5760}\delta^6_{-1}$
$+\frac{259}{5760}\delta^6_{0}$
$+\frac{259}{5760}\mu\delta^6_{\frac12}$
$+\frac{259}{5760}\delta^6_{1}$
$-\frac{341}{5760}\delta^6_{2}$
$+\frac{739}{5760}\delta^6_{3}$
$+\frac{12139}{5760}\delta^6_{4}$

$+\frac{25333}{11520}\delta^7_{-3\frac12}$
$+\frac{211}{2304}\delta^7_{-2\frac12}$
$-\frac{47}{1280}\delta^7_{-1\frac12}$
$+\frac{259}{11520}\delta^7_{-\frac12}$
$-\frac{259}{11520}\delta^7_{1\frac12}$
$+\frac{47}{1280}\delta^7_{2\frac12}$
$-\frac{211}{2304}\delta^7_{3\frac12}$
$-\frac{25333}{11520}\delta^7_{4\frac12}$

$+\frac{81227}{35840}\delta^8_{-4}$
$+\frac{21719}{3\,22560}\delta^8_{-3}$
$-\frac{869}{35840}\delta^8_{-2}$
$+\frac{447}{35840}\delta^8_{-1}$
$-\frac{3229}{3\,22560}\delta^8_{0}$
$-\frac{3229}{3\,22560}\mu\delta^8_{\frac12}$
$-\frac{3229}{3\,22560}\delta^8_{1}$
$+\frac{447}{35840}\delta^8_{2}$
$-\frac{869}{35840}\delta^8_{3}$
$+\frac{21719}{3\,22560}\delta^8_{4}$
$+\frac{81227}{35840}\delta^8_{5}$

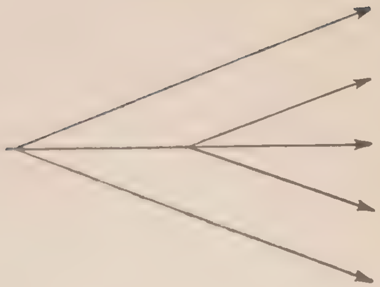
$+\frac{4\,98233}{2\,15040}\delta^9_{-4\frac12}$
$+\frac{1553}{30720}\delta^9_{-3\frac12}$
$-\frac{2165}{1\,29024}\delta^9_{-2\frac12}$
$+\frac{4817}{6\,45120}\delta^9_{-1\frac12}$
$-\frac{3229}{6\,45120}\delta^9_{-\frac12}$
$+\frac{3229}{6\,45120}\delta^9_{1\frac12}$
$-\frac{4817}{6\,45120}\delta^9_{2\frac12}$
$+\frac{2165}{1\,29024}\delta^9_{3\frac12}$
$-\frac{1553}{30720}\delta^9_{4\frac12}$
$-\frac{4\,98233}{2\,15040}\delta^9_{5\frac12}$

$-\frac{1215\,63409}{516\,09600}\delta^{10}_{-5}$
$+\frac{19\,87549}{516\,09600}\delta^{10}_{-4}$
$-\frac{6\,21491}{516\,09600}\delta^{10}_{-3}$
$-\frac{81503}{172\,03200}\delta^{10}_{-2}$
$-\frac{1\,40851}{516\,09600}\delta^{10}_{-1}$
$-\frac{1\,17469}{516\,09600}\delta^{10}_{0}$
$-\frac{1\,17469}{516\,09600}\mu\delta^{10}_{\frac12}$
$-\frac{1\,17469}{516\,09600}\delta^{10}_{1}$
$-\frac{1\,40851}{516\,09600}\delta^{10}_{2}$
$-\frac{81503}{172\,03200}\delta^{10}_{3}$
$-\frac{6\,21491}{516\,09600}\delta^{10}_{4}$
$+\frac{19\,87549}{516\,09600}\delta^{10}_{5}$
$-\frac{1215\,63409}{516\,09600}\delta^{10}_{6}$

$+\frac{2401\,83839}{1032\,19200}\delta^{11}_{-5\frac12}$
$+\frac{10\,18067}{344\,06400}\delta^{11}_{-4\frac12}$
$+\frac{1\,31171}{147\,45600}\delta^{11}_{-3\frac12}$
$+\frac{64057}{206\,43840}\delta^{11}_{-2\frac12}$
$+\frac{1\,64233}{1032\,19200}\delta^{11}_{-1\frac12}$
$+\frac{1\,17469}{1032\,19200}\delta^{11}_{-\frac12}$
$-\frac{1\,17469}{1032\,19200}\delta^{11}_{1\frac12}$
$-\frac{1\,64233}{1032\,19200}\delta^{11}_{2\frac12}$
$-\frac{64057}{206\,43840}\delta^{11}_{3\frac12}$
$-\frac{1\,31171}{147\,45600}\delta^{11}_{4\frac12}$
$-\frac{10\,18067}{344\,06400}\delta^{11}_{5\frac12}$
$-\frac{2401\,83839}{1032\,19200}\delta^{11}_{6\frac12}$

$+\frac{3\,28081\,17061}{1\,36249\,34400}\delta^{12}_{-6}$
$-\frac{3118\,51213}{1\,36249\,34400}\delta^{12}_{-5}$
$-\frac{916\,59719}{1\,36249\,34400}\delta^{12}_{-4}$
$-\frac{59\,68457}{27249\,86880}\delta^{12}_{-3}$
$-\frac{26\,65867}{27249\,86880}\delta^{12}_{-2}$
$-\frac{83\,40421}{1\,36249\,34400}\delta^{12}_{-1}$
$-\frac{71\,56487}{1\,36249\,34400}\delta^{12}_{0}$
$-\frac{71\,56487}{1\,36249\,34400}\mu\delta^{12}_{\frac12}$
$-\frac{71\,56487}{1\,36249\,34400}\delta^{12}_{1}$
$-\frac{83\,40421}{1\,36249\,34400}\delta^{12}_{2}$
$-\frac{26\,65867}{27249\,86880}\delta^{12}_{3}$
$-\frac{59\,68457}{27249\,86880}\delta^{12}_{4}$
$-\frac{916\,59719}{1\,36249\,34400}\delta^{12}_{5}$
$-\frac{3118\,51213}{1\,36249\,34400}\delta^{12}_{6}$
$+\frac{3\,28081\,17061}{1\,36249\,34400}\delta^{12}_{7}$

TABLE VI. 3. $D^3y_{\frac{1}{2}}$



									$+\frac{5005\,69373}{774\,14400}\delta^{12}_{-6}$
								$+\frac{3161\,11237}{516\,09600}\delta^{11}_{-5\frac{1}{2}}$	
							$+\frac{11\,11619}{1\,93536}\delta^{10}_{-5}$		$+\frac{105\,61007}{309\,65760}\delta^{12}_{-5}$
						$+\frac{51\,42611}{9\,67680}\delta^{9}_{-4\frac{1}{2}}$		$+\frac{84\,34073}{221\,18400}\delta^{11}_{-4\frac{1}{2}}$	
					$+\frac{2317}{480}\delta^{8}_{-4}$		$+\frac{1\,03871}{2\,41920}\delta^{10}_{-4}$		$-\frac{15\,58369}{387\,07200}\delta^{12}_{-4}$
				$+\frac{8197}{1920}\delta^{7}_{-3\frac{1}{2}}$		$+\frac{4\,71539}{9\,67680}\delta^{9}_{-3\frac{1}{2}}$		$-\frac{24\,79643}{516\,09600}\delta^{11}_{-3\frac{1}{2}}$	
			$+\frac{29}{8}\delta^{6}_{-3}$		$+\frac{357}{640}\delta^{8}_{-3}$		$-\frac{3737}{64512}\delta^{10}_{-3}$		$+\frac{12\,05453}{1548\,28800}\delta^{12}_{-3}$
		$+\frac{23}{8}\delta^{5}_{-2\frac{1}{2}}$		$+\frac{1237}{1920}\delta^{7}_{-2\frac{1}{2}}$		$-\frac{13649}{1\,93536}\delta^{9}_{-2\frac{1}{2}}$		$+\frac{72851}{73\,72800}\delta^{11}_{-2\frac{1}{2}}$	
	$+2\delta^{4}_{-2}$		$+\frac{3}{4}\delta^{6}_{-2}$		$-\frac{83}{960}\delta^{8}_{-2}$		$+\frac{1219}{96768}\delta^{10}_{-2}$		$-\frac{1\,62209}{774\,14400}\delta^{12}_{-2}$
$+1\delta^{3}_{-1\frac{1}{2}}$		$+\frac{7}{8}\delta^{5}_{-1\frac{1}{2}}$		$-\frac{203}{1920}\delta^{7}_{-1\frac{1}{2}}$		$+\frac{15419}{9\,67680}\delta^{9}_{-1\frac{1}{2}}$		$-\frac{4\,20529}{1548\,28800}\delta^{11}_{-1\frac{1}{2}}$	
	$-1\delta^{4}_{-1}$		$-\frac{1}{8}\delta^{6}_{-1}$		$+\frac{37}{1920}\delta^{8}_{-1}$		$-\frac{3229}{9\,67680}\delta^{10}_{-1}$		$+\frac{10679}{172\,03200}\delta^{12}_{-1}$
$+1\delta^{3}_{-\frac{1}{2}}$		$-\frac{1}{8}\delta^{5}_{-\frac{1}{2}}$		$+\frac{37}{1920}\delta^{7}_{-\frac{1}{2}}$		$-\frac{3229}{9\,67680}\delta^{9}_{-\frac{1}{2}}$		$+\frac{10679}{172\,03200}\delta^{11}_{-\frac{1}{2}}$	
$+\mathbf{1}\delta^{3}_{\frac{1}{2}}$		$-\mathbf{\frac{1}{8}}\delta^{5}_{\frac{1}{2}}$		$+\mathbf{\frac{37}{1920}}\delta^{7}_{\frac{1}{2}}$		$-\mathbf{\frac{3229}{9\,67680}}\delta^{9}_{\frac{1}{2}}$		$+\mathbf{\frac{10679}{172\,03200}}\delta^{11}_{\frac{1}{2}}$	
$+1\delta^{3}_{1\frac{1}{2}}$		$-\frac{1}{8}\delta^{5}_{1\frac{1}{2}}$		$+\frac{37}{1920}\delta^{7}_{1\frac{1}{2}}$		$-\frac{3229}{9\,67680}\delta^{9}_{1\frac{1}{2}}$		$+\frac{10679}{172\,03200}\delta^{11}_{1\frac{1}{2}}$	
	$-1\delta^{4}_{2}$		$+\frac{1}{8}\delta^{6}_{2}$		$-\frac{37}{1920}\delta^{8}_{2}$		$+\frac{3229}{9\,67680}\delta^{10}_{2}$		$-\frac{10679}{172\,03200}\delta^{12}_{2}$
$+1\delta^{3}_{2\frac{1}{2}}$		$+\frac{7}{8}\delta^{5}_{2\frac{1}{2}}$		$-\frac{203}{1920}\delta^{7}_{2\frac{1}{2}}$		$+\frac{15419}{9\,67680}\delta^{9}_{2\frac{1}{2}}$		$-\frac{4\,20529}{1548\,28800}\delta^{11}_{2\frac{1}{2}}$	
	$-2\delta^{4}_{3}$		$-\frac{3}{4}\delta^{6}_{3}$		$+\frac{83}{960}\delta^{8}_{3}$		$-\frac{1219}{96768}\delta^{10}_{3}$		$+\frac{1\,62209}{774\,14400}\delta^{12}_{3}$
		$+\frac{23}{8}\delta^{5}_{3\frac{1}{2}}$		$+\frac{1237}{1920}\delta^{7}_{3\frac{1}{2}}$		$-\frac{13649}{1\,93536}\delta^{9}_{3\frac{1}{2}}$		$+\frac{72851}{73\,72800}\delta^{11}_{3\frac{1}{2}}$	
			$-\frac{29}{8}\delta^{6}_{4}$		$-\frac{357}{640}\delta^{8}_{4}$		$+\frac{3737}{64512}\delta^{10}_{4}$		$-\frac{12\,05453}{1548\,28800}\delta^{12}_{4}$
				$+\frac{8197}{1920}\delta^{7}_{4\frac{1}{2}}$		$+\frac{4\,71539}{9\,67680}\delta^{9}_{4\frac{1}{2}}$		$-\frac{24\,79643}{516\,09600}\delta^{11}_{4\frac{1}{2}}$	
					$-\frac{2317}{480}\delta^{8}_{5}$		$-\frac{1\,03871}{2\,41920}\delta^{10}_{5}$		$+\frac{15\,58369}{387\,07200}\delta^{12}_{5}$
						$+\frac{51\,42611}{9\,67680}\delta^{9}_{5\frac{1}{2}}$		$+\frac{84\,34073}{221\,18400}\delta^{11}_{5\frac{1}{2}}$	
							$-\frac{11\,11619}{1\,93536}\delta^{10}_{6}$		$-\frac{105\,61007}{309\,65760}\delta^{12}_{6}$
								$+\frac{3161\,11237}{516\,09600}\delta^{11}_{6\frac{1}{2}}$	
									$-\frac{5005\,69373}{774\,14400}\delta^{12}_{7}$

II. *The Pressure Wave System in the Periodic Flow of Liquid in a Pipe.*

By R. S. Silver, M.A., B.Sc., Ph.D., Chief Physicist,
G. & J. Weir, Ltd., Engineers, Cathcart, Glasgow, S. 4.*

[Received December 9, 1940.]

Introduction.

Frequently in practical cases the flow of liquid in a pipe has a regularly repeated wave form. Where this is the case, the velocity of flow through any cross-section may be represented by a Fourier series, the independent variable being time. When an increment of velocity Δv occurs at one point, this is propagated through the liquid with the appropriate acoustic velocity and it is known that there is associated with it a pressure increment Δp which is proportional to Δv. From the theory of propagation of waves in a fluid (1) it is known that

$$\Delta p = r\Delta v \text{ where } r = \sqrt{\frac{k\rho}{g}},$$

where all quantities are in gravitational units. The quantity k is the bulk modulus of the fluid and ρ is its density.

In this discussion we shall confine ourselves to the case of a horizontal pipe of length L and of uniform cross-section S_1. Any particular problem in such a pipe will be specified in terms of the end conditions. We shall here consider that at one end the pipe receives a supply of liquid from a source, the supply being such that the velocity at the inlet may be represented in terms of a simple harmonic fluctuation about a mean value. At the other end we shall consider that the pipe adjoins another pipe of different cross-sectional area S_2, but of infinite length. Since any periodic flow can be represented by a Fourier series, the solution obtained for the simple harmonic case would lead directly to the general solution for any periodic flow by summation.

Our first problem is to consider what happens at the change of area. The arriving pulses will be in part transmitted through into the new pipe and in part reflected back into the first length. Since the following pipe is of infinite length, no disturbances will be reflected back from it. The laws of the reflexion and transmission of such pulses at a change of area have previously been derived by Diederichs and Pomeroy (2) by an approximate treatment. The following derivation, however, is somewhat more rigid and makes use of the Bernoulli equation. It shows that the laws obtained by Diederichs and Pomeroy are true as a first approximation in normal conditions.

---

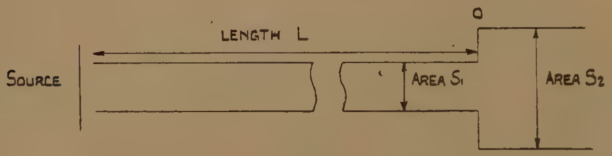
* Communicated by Professor E. Taylor Jones.

SECTION I.

Derivation of Laws of Reflexion and Transmission at Area Changes.

Referring to the figure, let us consider an instant at which the fluid velocity in the pipe just to the left of O is y, so that the velocity just to the right of O is $\frac{S_1}{S_2}y$. Then, if p_1 is the pressure just inside O at that instant, and p_2 the pressure in the passage just outside O, we have, by Bernoulli's theorem

$$p_1+\tfrac{1}{2}\frac{\rho}{g}y^2=p_2+\tfrac{1}{2}\frac{\rho}{g}\left(\frac{S_1}{S_2}\right)^2 y^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (1)$$



Now let the source supply a small increase in velocity of amount Δv. Then a wave is initiated from it which will cause an increase of y by an amount Δy. We require to determine Δy in terms of Δv and of the areas, *i. e.*, we require to find what change in velocity will be caused at O by the pulse Δv.

Along with the pulse Δv, there is also initiated a pressure pulse $r\Delta v$. Therefore, after the pulse has arrived, the pressure just to the left of O will be $p_1+r\Delta v+p_{z_1}$, where p_{z_1} is the reflected pressure wave which arrives by reflexion at O. We require also to determine its amount. Both y and p_{z_1} can be determined as follows :—

Apply Bernoulli's theorem to the new state just after the pulse has arrived. Since just on the right of O we must have initiated a velocity pulse of amount $\frac{S_1}{S_2}\Delta y$ and a corresponding pressure pulse of $\frac{rS_1}{S_2}\Delta y$, we have

$$p_1+r\Delta v+p_{z_1}+\tfrac{1}{2}\frac{\rho}{g}(y+\Delta y)^2=p_2+\frac{rS_1}{S_2}\Delta y+\tfrac{1}{2}\frac{\rho}{g}\left(\frac{S_1}{S_2}\right)^2(y+\Delta y)^2. \qquad (2)$$

Substituting from equation (1) in equation (2), and neglecting squares of Δy, we have

$$r\Delta v+p_{z_1}+\frac{\rho}{g}y\Delta y=\frac{rS_1}{S_2}\Delta y+\frac{\rho}{g}\left(\frac{S_1}{S_2}\right)^2 y\Delta y. \quad . \quad . \quad . \quad . \quad (3)$$

Now from elasticity theory we know also that the initial pulses Δv can be reduced to Δy only because of the effect of the reflected wave p_{z_1}, and that the reduction effected is proportional to p_{z_1}, *i. e.*, we have

$$\Delta v-\Delta y=\frac{1}{r}p_{z_1}.$$

$$\therefore \qquad p_{z_1}=r\Delta v-r\Delta y. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Substituting from equation (4) in equation (3) we have

$$2r\Delta v+r\Delta y\left(\frac{\rho}{rg}y-1\right)=r\Delta y\left[\frac{S_1}{S_2}+\frac{\rho}{rg}\left(\frac{S_1}{S_2}\right)^2 y\right].$$

Simplifying this gives

$$\Delta y=\frac{2S_2}{S_1+S_2}\frac{\Delta v}{\left[1+\frac{\rho}{rg}y\left(\frac{S_1}{S_2}-1\right)\right]}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Now $\qquad r=\sqrt{\frac{\rho k}{g}}, \quad \therefore \frac{\rho}{rg}=\sqrt{\frac{\rho}{gk}}=\frac{1}{c},$

where c is the velocity of sound in the fluid.

$$\therefore \qquad \Delta y=\frac{2S_2}{S_1+S_2}\frac{\Delta v}{\left[1+\frac{y}{c}\left(\frac{S_1}{S_2}-1\right)\right]}.$$

Now the velocity y is, in all practical cases of liquid flow, very much less than the velocity of sound in the liquid, and hence y/c is so small as to be negligible. We have, therefore, simply

$$\Delta y=\frac{2S_2}{S_1+S_2}(\Delta v). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Equation (6) shows how the velocity change caused at a change of area differs from the initiated pulse, by an amount depending on the areas. Equation (6) is the same as that derived by Diederichs and Pomeroy, and it is clear that it is true only because in practical cases y/c is so small as to be negligible.

Laws of Reflexion.

The reflected pressure wave p_{z_1} is then given by substituting equation (6) in equation (4), viz.,

$$p_{z_1}=r\Delta v\left(1-\frac{2S_2}{S_1+S_2}\right)=r\Delta v\frac{S_1-S_2}{S_1+S_2}.$$

∴ reflected pressure pulse

$$=\left(\frac{S_1-S_2}{S_1+S_2}\right) \text{(incident pressure pulse)}. \quad . \quad . \quad . \quad . \quad (7)$$

We can also regard it that the velocity is changed from Δv to Δy because of a reflected velocity pulse of amount $-(\Delta v-\Delta y)$. This reflected pulse will then be

$$-\Delta v\left(1-\frac{2S_2}{S_1+S_2}\right)=-\Delta v\left(\frac{S_1-S_2}{S_1+S_2}\right).$$

∴ reflected velocity pulse

$$=-\left(\frac{S_1-S_2}{S_1+S_2}\right) \text{(incident velocity pulse)}. \quad . \quad . \quad . \quad . \quad (8)$$

Equations (7) and (8) give the laws of reflexion of pressure and velocity pulses. They show that if $S_2<S_1$, *i. e.*, if the incident wave moves towards a constriction, the reflected pressure pulse has the same sign as the incident, while the reflected velocity pulse is opposite in sign to the incident. If, however, $S_2>S_1$, *i. e.*, if the incident wave moves towards an expansion, the reflected pressure pulse is of opposite sign, while the reflected velocity pulse is of the same sign.

It is convenient to define the quantity $\frac{S_1-S_2}{S_1+S_2}$, which is clearly a reflexion coefficient, as x_{12}.

Laws of Transmission.

The transmitted velocity pulse into the section S_2 is

$$\frac{S_1}{S_2}\Delta y=\frac{2S_1}{S_1+S_2}\Delta v=(1+x_{12})\Delta v.$$

∴ transmitted velocity pulse

$$=(1+x_{12}) \text{ (incident velocity pulse)},$$

and transmitted pressure pulse

$$=(1+x_{12})r\Delta v=(1+x_{12}) \text{ (incident pressure pulse)}.$$

Numerically, since x_{12}, even if negative, is necessarily less than 1, the transmitted pulses are clearly always of the same sign as the incident pulses.

SECTION II.

The Pressure Wave System in the Length L.

Let us denote the reflexion coefficient $\frac{S_1-S_2}{S_1+S_2}$ by x and let the velocity at the source end of the pipe be given by $q=q_m(1+h \sin m\omega t)$, where q_m

is the mean velocity and $q_m h$ is the amplitude of the fluctuations. The frequency of the fluctuations is $\dfrac{m\omega}{2\pi}$.

The velocity increment Δq starting from the source end at a time t will therefore be

$$\Delta q = (m\omega q_m h \cos m\omega t)\Delta t,$$

and the associated pressure pulse

$$\Delta p = (rm\omega q_m h \cos m\omega t)\Delta t.$$

Consider now a point at distance y from the source end of the pipe. Then at an instant t there will be arriving at this point the following pressure pulses :—

(1) that which left the source end at the time $t-\dfrac{y}{c}$;

(2) that which left the source end at the time $t-\dfrac{2L}{c}+\dfrac{y}{c}$ and suffered one reflexion at O ;

(3) that which left the source end at a time $t-\dfrac{2L}{c}-\dfrac{y}{c}$ and suffered one reflexion at O and one at the source end ;

(4) that which left the source end at a time $t-\dfrac{4L}{c}+\dfrac{y}{c}$, was reflected at O, back again from the source end, and back again from O ;

(5) and so on, to infinity.

We shall assume that reflexion from the source end is equivalent to reflexion from a closed end, *i. e.*, there is no transmission and reflexion is complete, so that for these reflexions the coefficient is 1.

The above impulses will, therefore, be in magnitude as follows :—

$$(1)\quad \left[rm\omega q_m h \cos m\omega\left(t-\frac{y}{c}\right)\right]\Delta t;$$

$$(2)\quad x\left[rm\omega q_m h \cos m\omega\left(t-\frac{2L}{c}+\frac{y}{c}\right)\right]\Delta t;$$

$$(3)\quad x\left[rm\omega q_m h \cos m\omega\left(t-\frac{2L}{c}-\frac{y}{c}\right)\right]\Delta t;$$

etc.

Hence for the actual rate of pressure change at the point we shall have to sum the infinite series and obtain

$$\frac{dp}{dt} = rm\omega q_m h\left\{\cos m\omega\left(t-\frac{y}{c}\right)+x\cos m\omega\left(t-\frac{2L}{c}+\frac{y}{c}\right)\right.$$

$$+x\cos m\omega\left(t-\frac{2L}{c}-\frac{y}{c}\right)+x^2\cos m\omega\left(t-\frac{4L}{c}+\frac{y}{c}\right)$$
$$+x^2\cos m\omega\left(t-\frac{4L}{c}-\frac{y}{c}\right)+\ldots\text{etc.}\Big\}.$$

Since x is less than 1 it can be proved that the series is convergent, and indeed its sum can be obtained. We shall, however, integrate first term by term and sum instead the resultant series.

Integrating, we find for the pressure p at any time t at any point distant y from the source end:

$$\begin{aligned}p-p_{\text{mean}}=rq_m h\Big[&\sin m\omega\left(t-\frac{y}{c}\right)\\&+x\sin m\omega\left(t-\frac{2L}{c}+\frac{y}{c}\right)\\&+x\sin m\omega\left(t-\frac{2L}{c}-\frac{y}{c}\right)\\&+x^2\sin m\omega\left(t-\frac{4L}{c}+\frac{y}{c}\right)\\&+x^2\sin m\omega\left(t-\frac{4L}{c}-\frac{y}{c}\right)+\ldots\text{etc.}\Big].\end{aligned}$$

Now, for convenience, write

$$\frac{2L}{c}=\mu,$$

$$m\omega\mu=\alpha,$$

$$m\omega\frac{y}{c}=\beta.$$

$$\begin{aligned}\therefore\quad p-p_{\text{mean}}=rq_m h[&\sin(m\omega t-\beta)\\&+x\sin(m\omega t-\alpha+\beta)+x\sin(m\omega t-\alpha-\beta)\\&+x^2\sin(m\omega t-2\alpha+\beta)\ldots\text{etc.}].\end{aligned}$$

Now we have

$$\sin(m\omega t-S\alpha+\beta)=\sin(m\omega t+\beta)\cos S\alpha-\cos(m\omega t+\beta)\sin S\alpha.$$

$$\begin{aligned}\therefore\ p-p_{\text{mean}}&\\=rq_m h\{&\sin(m\omega t-\beta)[1+x\cos\alpha+x^2\cos 2\alpha+x^3\cos 3\alpha+\ldots\text{etc.}]\\&+\sin(m\omega t+\beta)[x\cos\alpha+x^2\cos 2\alpha+x^3\cos 3\alpha+\ldots\text{etc.}]\\&-\cos(m\omega t-\beta)[x\sin\alpha+x^2\sin 2\alpha+x^3\sin 3\alpha+\ldots\text{etc.}]\\&-\cos(m\omega t+\beta)[x\sin\alpha+x^2\sin 2\alpha+x^3\sin 3\alpha+\ldots\text{etc.}]\}.\end{aligned}$$

Now, by considering $z=\cos\theta+i\sin\theta=e^{i\theta}$ and using the expression for $\dfrac{1}{1-xz}$, it can be shown that

$$1+x\cos\alpha+x^2\cos 2\alpha+\ldots\text{etc.}=\frac{1-x\cos\alpha}{1-2x\cos\alpha+x^2}$$

and

$$x\sin\alpha+x^2\sin 2\alpha+\ldots\text{etc.}=\frac{-x\sin\alpha}{1-2x\cos\alpha+x^2}.$$

$$\therefore\ p-p_{\text{mean}}=rq_m h\Bigg[\sin(m\omega t-\beta)+\sin(m\omega t+\beta)\left(\frac{1-x\cos\alpha}{1-2\,x\cos\alpha+x^2}\right)$$
$$+(\cos m\omega t-\beta)+\cos(m\omega t+\beta)\left(\frac{x\sin\alpha}{1-2x\cos\alpha+x^2}\right)-\sin(m\omega t+\beta)\Bigg].$$

Now

$$\sin A+\sin B=2\sin\frac{A+B}{2}\cos\frac{A-B}{2};\ \cos A+\cos B=2\cos\frac{A+B}{2}\cos\frac{A-B}{2}.$$

$$\therefore\ p-p_{\text{mean}}=rq_m h\Bigg\{2\left(\frac{1-x\cos\alpha}{1-2x\cos\alpha+x^2}\right)\sin m\omega t\cos\beta$$
$$+\frac{2x\sin\alpha}{1-2\,x\cos\alpha+x^2}\cos m\omega t\cos\beta$$
$$-\sin m\omega t\cos\beta-\cos m\omega t\sin\beta\Bigg\}$$
$$=rq_m h\Bigg\{\sin m\omega t\left[\frac{2(1-x\cos\alpha)\cos\beta}{1-2x\cos\alpha+x^2}-\cos\beta\right]$$
$$+\cos m\omega t\left[\frac{2x\sin\alpha\cos\beta}{1-2x\cos\alpha+x^2}-\sin\beta\right]\Bigg\}$$
$$=rq_m h\Bigg\{\sin m\omega t\left[\frac{\cos\beta(1-x^2)}{1-2x\cos\alpha+x^2}\right]$$
$$+\cos m\omega t\left[\frac{2x\sin(\alpha+\beta)-\sin\beta(1-x^2)}{1-2x\cos\alpha+x^2}\right]\Bigg\}.$$

Now write the coefficient of $\sin m\omega t$ as a_y and the coefficient of $\cos m\omega t$ as b_y.

$$\therefore\ p-p_{\text{mean}}=rq_m h[a_y\sin m\omega t+b_y\cos m\omega t]$$
$$=rq_m h\sqrt{a_y^2+b_y^2}\,[\sin(m\omega t+\theta)],\quad . \quad (9)$$

where θ is an angle whose tangent is b_y/a_y.

Equation (9) shows that, in these circumstances, the amplitude of the pressure variation at the point y is

$$rq_m h\sqrt{a_y^2+b_y^2},$$

and its frequency corresponds to that of the source.

Hence the maximum and minimum amplitudes are determined by the maximum and minimum amplitudes of $\sqrt{a_y^2+b_y^2}$, which must now be considered.

Expanding and simplifying

$$\sqrt{a_y^2+b_y^2}=\sqrt{\frac{(1-x^2)+4x^2\sin^2(\alpha+\beta)-4x(1+x^2)\sin\beta\sin(\alpha+\beta)}{1-2x\cos\alpha+x^2}}.$$

We first consider the maximum and minimum values of this expression with respect to β, *i. e.*, we consider the positions in the pipe at which maxima and minima occur. There are turning points if

$$\frac{d}{d\beta}\sqrt{a_y^2+b_y^2}=0.$$

By differentiating we find that this occurs when

$$\tan 2\beta=\frac{(1+x^2)\sin\alpha-x\sin 2\alpha}{x\cos 2\alpha-(1+x^2)\cos\alpha}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Equation (10) determines the positions in the pipe at which for any given value of α, *i. e.*, for any source frequency, the surge is greatest and least, *i. e.*, it determines the nodes and anti-nodes of the complex vibration, *e. g.*, if $\alpha=2s\pi$ where s is any integer

$$\tan 2\beta=0.$$

$$\therefore \qquad 2\beta=0,\pi,2\pi,3\pi,\text{ etc.};\quad\therefore\ \beta=0,\ \frac{\pi}{2},\ \pi,\ \frac{3\pi}{2},\text{ etc.}$$

But $$\alpha=m\omega\frac{2L}{c}\quad\text{and}\quad\beta=m\omega\frac{y}{c};\quad\therefore\ y=2L\frac{\beta}{\alpha}.$$

$$\therefore \qquad y=0,\ \frac{L}{2s},\ \frac{L}{s},\ \frac{3L}{2s},\ \frac{2L}{s},\text{ etc.}$$

Testing these in equation for $\sqrt{a_y^2+b_y^2}$, we see that if x is positive $\sqrt{a_y^2+b_y^2}$ is a maximum at $y=0,\ \frac{L}{s},\ \frac{2L}{s}$, etc. and a minimum at $y=\frac{L}{2s},\ \frac{3L}{2s}$, etc. Whereas if x is negative $y=0,\ \frac{L}{s},\ \frac{2L}{s}$, etc. are minimum points, and the maxima appear at $y=\frac{L}{2s},\ \frac{3L}{2s}$, etc.

On the other hand, if $\alpha=(2s+1)\pi$, if x is positive, the points

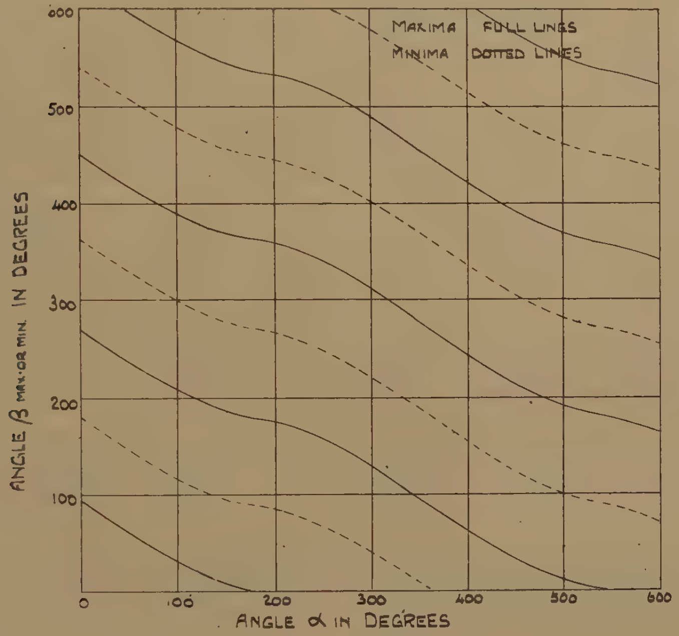
$$y=\frac{L}{2s+1},\ \frac{3L}{2s+1},\ \frac{5L}{2s+1},\text{ etc.}$$

are maxima and

$$0,\ \frac{2L}{2s+1},\ \frac{4L}{2s+1},\text{ etc.}$$

are minima, while if x is negative, the reverse is the case.

Apart from these particular values of α, it is interesting to consider values of β which will satisfy equation (10) for any values of α. The solution is best represented by plotting the results as in the graph. In the graph the angle β is plotted against the angle α and separate lines occur for each solution. Since, for a given length of pipe α is proportional to the frequency of the source, and for a given source frequency β is proportional to the distance from the source, the graph shows very elegantly how, as the frequency increases, the points of maximum and minimum disturbance become displaced towards the source as well as being bunched together with shorter wave-length. When the frequency diminishes the wave system slides along the pipe away from the source, and as one point goes out at the far end another comes in from the source.



APPENDIX I.

The Reflexion Conditions at an Orifice.

It is of interest to derive the reflexion conditions at an orifice through which fluid is being forced. This may be done by an extension of the method in which Bernoulli's theorem was used to derive the general reflexion coefficients. At any instant the equation for the conditions just before the orifice and in the orifice passage is

$$p+\tfrac{1}{2}\frac{\rho}{g}q^2=p_{\text{ex}}+\tfrac{1}{2}\frac{\rho}{g}\left(\frac{A}{a}\right)q^2,$$

where A is the cross-sectional area before the orifice and a is the cross-sectional area of the aperture, while p_{ex} is the pressure in the aperture.

Now consider the arrival of a velocity pulse Δq accompanied by a pressure pulse $r\Delta q$ and let the reflected wave of pressure be p_z. Then, in the instant immediately after reflexion, we shall have

$$p_z+p+r\Delta q+\tfrac{1}{2}\frac{\rho}{g}(q+\Delta y)^2=p_{ex}+\tfrac{1}{2}\frac{\rho}{g}\left(\frac{A}{a}\right)^2(q+y)^2+r\frac{A}{a}\Delta y,$$

where Δy is the net velocity pulse after reflexion. Now since the condition is that of an orifice opening to constant pressure, the term $r\frac{A}{a}\Delta y$ on the right-hand side is immediately cancelled and, therefore, in the correct reflexion equation, must be omitted. When this is done and the values substituted from the first equation into the second equation, we get

$$p_z+r\Delta q+\frac{\rho}{g}q\Delta y=\frac{\rho}{g}\left(\frac{A}{a}\right)^2 q\Delta y.$$

But $p_z=r(\Delta q-\Delta y)$. Substitute

$$\Delta y=\frac{2\Delta q}{1+\frac{\rho}{rg}\left(\frac{A}{a}\right)^2-1q}.$$

Since q varies, we shall have to use the mean value of q_m for q in this equation. We also have, if p_1 is the mean pressure difference across the aperture,

$$p_1=\tfrac{1}{2}\frac{\rho}{g}q_m^2\left[\left(\frac{A}{a}\right)^2-1\right],\quad\text{and hence}\quad\frac{\rho}{rg}\left[\left(\frac{A}{a}\right)^2-1\right],$$

$$q_m=\frac{2p_1}{rq_m}.$$

We therefore obtain

$$\Delta y=\frac{2\Delta q}{1+\frac{2p_1}{rq_m}}.$$

Thus

$$p_z=r\Delta q-r\Delta y=r\Delta q\left(1-\frac{2}{1+\frac{2p_1}{rq_m}}\right)$$

$$=r\Delta q\left(\frac{\frac{2p_1}{rq_m}-1}{\frac{2p_1}{rq_m}+1}\right).$$

But, by definition of reflexion, coefficient $p_2 = xr\Delta q$.

$$\therefore \qquad x = \frac{\dfrac{2p_1}{rq_m} - 1}{\dfrac{2p_1}{rq_m} + 1}.$$

This last equation therefore gives the value of the reflexion coefficient at an orifice through which fluid is being forced to a region of constant pressure.

In conclusion, I wish to express my gratitude to Messrs. G. & J. Weir, Limited, Cathcart, Glasgow, for permission to publish this paper, which forms part of a programme of experimental and theoretical research carried out in their Research Department.

References.

(1) 'The Mechanical Properties of Fluids,' Blackie & Son, 2nd edition, p. 223 (1936).
(2) Diederichs and Pomeroy, Trans. A. S. M. E., Petroleum Section. p. 46 (May–Aug. 1939).

Research Dept.,
G. & J. Weir, Ltd.,
Glasgow.

III. *Dimensions of Physical Quantities.*

By Professor W. Wilson, F.R.S.[*]

[Received November 10, 1941.]

On the first page of his 'Treatise on Electricity and Magnetism,' after some preliminary explanations about the Unit and Numerical Value and the derivation of units in dynamical sciences from three fundamental units of Length, Time and Mass, Clerk Maxwell says " The formulæ at which we arrive must be such that a person of any nation, by substituting for the different symbols the numerical values of the quantities as measured by his own national units, would arrive at a true result." He is thus led to the following definition of the *dimensions* of a *unit*: " When a given unit varies as the nth power of one of these units," *i. e.*, of one of the fundamental units, " it is said to be of n dimensions as regards that unit." Thus the equation which defines the conventional measure of a force may be written

$$f[\mathrm{F}]=[\mathrm{M}][\mathrm{L}][\mathrm{T}^{-2}]m\frac{d^2l}{dt^2}, \quad \ldots \ldots \ldots \quad (1)$$

where [F], [M], etc., indicate the magnitudes of the corresponding unit. If the unit of mass were multiplied by α, that of length by β and that of time by γ, the unit of force would obviously be multiplied by

$$\alpha\beta\gamma^{-2},$$

and the corresponding numerical measure divided by this number. Similarly, if the numerical measures of mass, length and time were multiplied by α, β and γ respectively, that of the force would be multiplied by

$$\alpha\beta\gamma^{-2},$$

and that of the unit of force would be divided by this number.

Maxwell speaks of the dimensions of a unit, and the abbreviated statement of the dimensions of the unit of force is

$$\mathrm{F}=\mathrm{MLT}^{-2}. \quad \ldots \ldots \ldots \quad (2)$$

The reciprocal relationship, already described, between the numerical quantity and the associated unit justifies the use of the attribute *dimensions* for the physical quantity itself, and I prefer to speak of the

---

[*] Communicated by the Author.

dimensions of a physical quantity, partly because there are important questions which arise in the definition of different derived units and are concerned with dimensionless quantities which naturally do not appear in statements of dimensions.

For example, one might adopt as the unit of volume the volume of a sphere of unit radius. This choice of a unit may be unpractical and inconvenient, but it is certainly permissible. Thus the volume of a sphere may be expressed by

$$V=R^3, \qquad (3)$$

and in terms of this unit the volume of a rectangular parallelopipedon is expressed by

$$V=\frac{3}{4\pi}L\,.\,B\,.\,H, \qquad (4)$$

where L, B, and H are respectively its length, breadth and height. More generally (4) might be written

$$V=\alpha\,L\,.\,B\,.\,H, \qquad (5)$$

where α is dimensionless—I hesitate to use the term " pure number."

This dimensionless quantity is of great importance in electromagnetism. The formula for the force between two point charges may (indeed, I think it should) be written

$$F=AQ_1Q_2/Kr^2, \qquad (6)$$

The dimensionless quantity, A, is usually unity, but Lorentz and Heaviside preferred

$$A=1/4\pi.$$

It seems almost unnecessary to draw the following inference from (6) combined with (2) about the dimensions of an electric charge :

$$Q=M^{1/2}L^{3/2}T^{-1}K^{1/2}. \qquad (7)$$

Every electromagnetic relationship contains at least *two* electric or magnetic quantities. Thus in (6) we have Q and K. We can only eliminate one of these by introducing some new electric or magnetic quantity. This explains why one cannot express the dimensions of Q in terms of those of L, M, and T only.

Three important field equations are included in the statement

$$\beta \boldsymbol{i}=\text{curl }\mathbf{H}, \qquad (8)$$

where $\boldsymbol{i}$ means current density (and many include, of course, displacement current density) and **H** means magnetic field intensity. The quantity β is another dimensionless quantity whose rôle is like that of α in (5) or that of A in (6). From (8) we obtain the dimensional equation

$$Q=M^{1/2}L^{1/2}\mu^{-1/2}. \qquad (9)$$

The definition of the meaning of "dimensions," which I have adopted from Clerk Maxwell himself, assigns *the same dimensions* to the Q in (7) and the Q in (9). Hence

$$M^{1/2}L^{3/2}T^{-1}K^{1/2}=M^{1/2}L^{1/2}\mu^{-1/2}$$

or

$$K^{-1/2}\mu^{-1/2}=LT^{-2}. \quad \text{. . . . .} \quad (10)$$

In Clerk Maxwell's time, and indeed long after, it was generally, and mistakenly, believed that all physical phenomena were, *au fond*, mechanical, and that consequently Q, K, μ, etc., should have dimensions expressible in terms of those of one or more of length, mass and time. The only way of doing this is to suppress the dimensions of some electric or magnetic quantity. This means that in changing our units we restrict our freedom of action by artificially keeping the measure of some one electric or magnetic quantity unchanged. One method adopted was to suppress those dimensions of K in (7), another to suppress those of μ in (9). The former was associated with the choice of unity for the K of free space, the latter with the choice of unity for the μ of free space, and so the dimensions of Q were described as $M^{1/2}L^{3/2}T^{-1}$ "in the electrostatic system" and as $M^{1/2}L^{1/2}$ "in the electromagnetic system," and the ratio of Q (as measured in the E.S. system) to Q (as measured in the E.M. system), was given the dimensions of a velocity

$$\frac{Qe.s.}{Qe.m.}=LT^{-1}. \quad \text{. . . . .} \quad (11)$$

But (11) can only be regarded as a dimensional equation if the term "dimensions" is defined in a different way, and in a way which is incompatible with the definition of dimensions which has the sanction of Maxwell, and which in fact is the accepted one.

One might, for example, take a mass equal to that of a cubic centimetre of water at 4° C. and choose for our unit of force its weight, and it would be perfectly permissible in such a case to adopt the equation

$$1=mg. \quad \text{. . . . .} \quad (12)$$

The ratio of the unit of mass, thus defined, to the gramme would be g, or about 980·6 cm. sec.$^{-2}$.

It is hardly necessary to add that we are here using "dimensions" in a sense different from the Maxwellian one. Both of these discrepant senses have found their way into discussions on dimensions. It is claimed by G. B. Brown *, for example, that c in the equation

$$u=\frac{c}{\sqrt{\mu K}} \quad \text{. . . . .} \quad (13)$$

---

* G. B. Brown, "A New Treatment of the Theory of Dimensions," Proc. Phys. Soc. vol. liii. p. 418 (1941).

has the dimensions of a velocity. It can only have such dimensions in the sense of equation (11), and, of course, if we assign such dimensions to c in (13), the product μK must be dimensionless. But this is surely nonsense. We have only to think of such a formula as

$$T = 2\pi\sqrt{\lambda C}, \quad \ldots \ldots \ldots \quad (14)$$

where λ is the self-inductance in a circuit and C the capacitance (as it is now called). The dimensions of λ and C are

$$\lambda = \mu L,$$

$$C = KL,$$

and therefore

$$T = \sqrt{\mu K}\,L.$$

Now how is it possible to reconcile this with Brown's view that K and μ are "pure numbers"? Let us examine Brown's argument. He says that K (for example) is the "ratio of numbers obtained by similar measurements." But so is a length, which Brown does not regard as a "pure number." I might measure the length of a table and find it to be 394 inches, and by a precisely "similar" measurement I might determine the length of a metre to be 39·4 inches; but neither Dr. Brown nor I regard 10 metres as dimensionless. The corresponding "similar measurements" in the case of the di-electric constant of, say, sulphur are the capacity of a condenser with the sulphur between its plates and of the same condenser with its plates separated by free space. The second measurement in both cases, that of the lengths and that of the capacities, is quite independent of the first. To see how this differs from a "pure number" we may take the example of an angle. Here we divide the length of an arc by that of the *corresponding* radius. These two are not independent, one being proportional to the other.

The dimensionless character of "c" in (13) can be brought out in another way. Let us turn to the dimensionless quantities β, in (8), and A, in (6). The former is often chosen to be 4π (electromagnetic or electrostatic units of the old type). Sometimes it is unity (E.M. or E.S. units of the Lorentz–Heaviside type); but whatever choice we make for A in (6) and in the corresponding formula

$$F = Am_1m_2/\mu r^2,$$

for the force between magnetic point poles, we obtain for the velocity of E.M. waves in free space,

$$u = \frac{4\pi A/\beta}{\sqrt{\mu K}},$$

and the dimensionless character of the numerator is again evident. That it is sometimes chosen to be identical with "c" does not give it dimensions. The purpose of this choice is to enable us to take the dielectric constant and permeability of free space to be unity.

In thermal relationships temperature, for example, is *always* associated with another thermal quantity—specific heat or entropy. The former is not a dimensionless quantity. It has, in fact, the dimensions

$$\text{Energy}\times M^{-1}\times\theta^{-1},$$

where θ means temperature. One cannot (for the same reasons as have been given in connexion with quantity of electricity, dielectric constant, etc.) express the dimensions of temperature, specific heat, entropy, etc., in terms of those of one or more of length, mass and time.

It thus appears that we need one or more of *five* quantities, possessing the kind of independence that may be called dimensional, in order to express the dimensions of any other physical quantity. No more than five appear to be needed. These might be chosen to be

$$L, M, T, K, \text{ and } \theta,$$

and we need five so-called fundamental units, which might be the centimetre, gramme, mean solar second, dielectric constant of free space, and the centigrade degree of Kelvin's work scale. The number 5 would seem to be as important (perhaps more important) than Eddington's 4. It appears in Kaluza's 5-dimensional relativity theory, in which the conservations of mass, energy, electric charge, etc. (referred to by Brown), are all included in a generalized principle of conservation of momentum. This 5-fold multiplicity, which appears in equations representing physical relations, has, strangely enough, been known for a long time. Brown refers to Rücker's suggestion (a perfectly sound one) that we should adopt as fundamental units those of length, mass, time, magnetic permeability and temperature, and also to Tolman's equally sound suggestion that we should adopt length, mass, time, electric charge and entropy; but he does not appear to have inquired carefully into the reasons which moved Rücker and Tolman.

Let us agree with Brown that length and time are very important quantities, and in one special sense perhaps more fundamental than any other quantities. How can we (remembering what has already been said about dimensions and about *five* primary or fundamental quantities: length, mass, time and two others) eliminate from dimensional statements ,everything except length and time? We can only do this by making use of things established experimentally or by the adoption of principles or conventions which the theory indicates to be important, or finally by some artificiality which "suppresses" the dimensions of one or more quantities, as explained above and as illustrated by the assignment of the dimensions $M^{1/2}L^{3/2}T^{-1}$ to Q.

There is no vestige of doubt about the importance or fundamental character of Planck's constant h:

$$h=(\text{a momentum})\times(\text{a distance});$$

and it is obvious we may, if we choose, adopt the quotient of *mass/h* to serve as a substitute for mass. This would cause h to be unity, but is not a mere artificiality. If we represent this substitute mass by M′, its dimensions are

$$M'=TL^{-2}, \quad \ldots\ldots\ldots \quad (15)$$

and if we denote the measure of energy, which emerges from this convention, by E′, its dimensions are

$$E'=M'L^2T^{-2}$$

or

$$E'=T^{-1}, \quad \ldots\ldots\ldots \quad (16)$$

while those of momentum become

$$M'LT^{-1}=L^{-1}. \quad \ldots\ldots\ldots \quad (17)$$

These are obviously quite important and suggestive expressions. In the analogy discovered by Hamilton between geometrical optics and mechanics, the optical analogue of energy is a reciprocal time (frequency) while that of momentum is a reciprocal distance (wave-number). Maupertius' principle of stationary action, namely

$$\delta\int_1^2 p\,dq=0,$$

is the exact analogue of Fermat's principle of the stationary optical path, namely

$$\delta\int_1^2\frac{dq}{\lambda}=0.$$

The importance of these things is rendered the more evident when we remember that wave-mechanics emerges from the extension of this analogy beyond the limits of geometrical optics into that of undulatory optics *.

Let us compare the convention which introduces the mass M′ with another sort of mass substitute. We adopt the Newtonian formula

$$F=GM_1M_2/r^2, \quad \ldots\ldots\ldots \quad (18)$$

where G is the gravitational constant, and introduce a new quantity, M″, to serve the purposes of mass, defined by

$$M''=GM; \quad \ldots\ldots\ldots \quad (19)$$

---

* W. Wilson, 'Science Progress' (October 1937).

we thus find, if we define the measure of force by $d(M''v)/dt$, for the dimensions of M'',

$$M''=L^3T^{-2}. \qquad (20)$$

This mass, or substitute for mass, called gravitational mass, is referred to by Clerk Maxwell in his 'Treatise,' and Dr. Brown attaches importance to it. Is it really so important ? or is it in any sense of fundamental significance ? Clearly there are more conventions than one which enable us to define a mass substitute whose dimensions are derivable from length and time. I find nothing to say in favour of the gravitational substitute except its practical convenience for astronomers. Indeed, the equation (18), on which it rests, is in fact only an approximation to the truth. There is apparently no *unique* way of expressing the dimensions of mass in terms of those of length and time, and the expression (15) has certainly a closer association with fundamentals than (20), which Brown favours.

We may, of course, go still further and introduce as a time substitute the distance E.M. waves travel through free space during the time in question. Velocities will now become dimensionless, and the gravitational mass acquires the dimensions of a *length*. This is the justification of Eddington's description of the sun's mass as 1·47 kilometres. On the other hand, when we introduce this time substitute into the action mass M′ it acquires the dimensions of a *reciprocal length* ! With this convention Brown's "equivalent" dimensions reduce, in the case of energy, momentum and mass, simply to lengths ; but reference to (15), (16) and (17) shows that there is just as much (or perhaps more) reason for regarding them as *reciprocal* lengths.

Turning to thermal quantities, we may use as a substitute for temperature the Willard Gibbsian modulus which is equal to $k\theta$, where k is Boltzmann's constant and θ is the Kelvin work scale temperature. If we represent this temperature substitute by θ' its dimensions are those of energy. Entropy would thus acquire the dimensions of a "pure number," and since its nature appears to be that of a probability this would seem to be very appropriate.

We can go further still in the choice of substitutes for important physical quantities. There is as much, or more, to be said in favour of treating quantity of electricity as dimensionless than there is in favour of treating K and μ in this way. The fact that Q is distributed (or exists) in integral multiples of $4{\cdot}8\times10^{-10}$ E.S. units suggests the replacement of Q by a substitute, Q′, defined as the (algebraic) number of positrons and electrons which can carry the charge Q. Adopting the mass M′, the charge substitute Q′, which of course is dimensionless, and the time substitute described above, we can express the dimensions of all quantities in terms of length.

K and μ now become dimensionless, and we obtain the following interesting dimensional statements :—

Mass $=L^{-1}$,
Force $=L^{-2}$,
Electric field intensity $=L^{-2}$,
Electric displacement $=L^{-2}$,
Magnetic field intensity $=L^{-2}$,
Magnetic induction $=L^{-2}$,
Gravitational constant $=L^{2}$,
Electrical resistance $=L^{0}$,
Capacity and Inductance $=L^{1}$.

There are all kinds of things that can be done in this way, and though "the only measurements that physicists actually require to make are either of space-like or time-like intervals," there is no unique way of carying out "dimensional reasoning" in terms of only L and T. So Brown's theory (and indeed any other which attempts a unique statement for the dimensions of all physical quantities in terms of L and T only) fails.

IV. *The Steady Flow of Ground-Water to a Pumped Well in the Vicinity of a River.*

By N. S. Boulton, M.Sc., Assoc.M.Inst.C.E., University of Sheffield.

[Received September 17, 1941.]

Introduction.

The form and extent of the lowering in the water-table, or "cone of depression," produced by pumping from a well, penetrating a homogeneous water-bearing medium, is a question of practical importance and theoretical interest. The original formula, due to Dupuit, for calculating the yield to a well in relation to the draw-down, and the lowering of the ground-water surface around it, has often been reproduced in hydrological literature; but as writers have recently pointed out [†], the formula assumes the existence of a discontinuity in the water-bearing medium at a given distance from the well—a condition rarely, if ever, encountered in practice.

The rainfall seeping down to and feeding the water-table, more or less uniformly over its whole surface, is ultimately discharged at some area where the water-table intersects the ground-surface. Neglecting seasonal fluctuations, this is a flow-system in equilibrium. When pumping from a well is commenced the equilibrium is disturbed, and it is only re-established when the "cone of depression" due to pumping has extended to one or more areas of discharge, usually springs or seepages, such that the volume of discharge water is lessened by an amount equal to the steady rate of pumping. If the cone of depression extends to a surface stream, its flow may likewise contribute to the re-establishment of equilibrium.

The areas of contact between ground and surface flow thus have a stabilizing effect in making possible the re-establishment of true equilibrium. But only if these areas are in the immediate vicinity of the well will equilibrium become established within a reasonably short time after pumping commences.

If the natural water-table has no intersection with the ground-surface, the "cone" will never reach true equilibrium, but will continually

---

[*] Communicated by the Author.

[†] J. H. Steegewentz and B. A. Van Nes, 'Water,' vol. xli. no. 517, p. 561 (December 1939); O. E. Meinzer and L. K. Wenzel, 'Economic Geology,' vol. xxxv. no. 8, p. 926.

expand, though at a diminishing rate. This fact has recently been emphasized by Theis * who has given a formula for the lowering of the piezometric surface when non-equilibrium conditions exist in an aquifer bounded above as well as below by an impervious layer (artesian conditions).

Statement of Problem.

The problem considered in the present paper is to determine a general equation for the motion to a pumped well and to investigate the form and extent of the " cone of depression " under the following conditions :—

(*a*) The water-bearing medium is hydrologically isotropic. It is bounded above by the ground-surface and below by a horizontal impervious bed. The medium extends laterally a great (theoretically infinite) distance in all directions. The presence of fissures is thus excluded.

(*b*) There is a flowing river, with its bed unpuddled, having a straight course in the vicinity of the well.

(*c*) The well penetrates the permeable medium to its base and is unlined.

(*d*) The straight course of the river is (theoretically) infinite in length. The form of the river-bed is a half-circular cylinder with its axis horizontal. Other forms of cross-section may be included by estimating the radius of an " equivalent " semi-circular cross-section.

In practice the river-bed would have a small longitudinal gradient which produces the surface-water flow along it. This gradient is, however, neglected in the analysis.

(*e*) The radius of the cross-section of the river is small, both as compared with the thickness of the permeable medium and the distance of the well from the river.

(*f*) The river is assumed to flow full to the axis of the cylinder. The water-level will actually be lowered due to the reduction of surface-flow in the river caused by pumping. Simple hydraulic calculation shows, however, that this lowering is small compared with the differences of level in the water-table produced by pumping, provided that the river-flow is large compared with the quantity pumped. The lowering in river-level may then be neglected.

(*g*) Before pumping is commenced, the natural water-table is horizontal and at the level of the river-surface. From this condition, it follows that there is no initial ground-water flow to or from the river.

---

* 'Economic Geology,' vol. xxxiii. pp. 892, 895 (1938).

The only intersection of the water-table and ground-surface occurs at the river.

The gradient of the water-table fed by percolation from the rainfall, either towards the river or towards a distant outfall, is not considered in the present investigation. This simplification does not decrease the value of the results obtained since inclusion of the flow-system due to the rainfall does not appreciably affect the lowering of the water-table from the initial rest-level due to pumping.

(*h*) The well is pumped at a constant rate until equilibrium conditions exist (steady flow in the hydrodynamical sense). The total rate of flow underground from the river then equals the rate of pumping.

(*i*) The flow in the permeable medium obeys Darcy's law.

(*j*) The inertia forces involved may be neglected *.

(*k*) The gravity-flow system with a free surface, which has been defined, cannot be solved by known methods. An approximate solution may, however, be obtained by considering a slightly modified flow-system without a free surface in the permeable medium. For purposes of analysis, a horizontal impervious layer is introduced into the permeable medium at river-level. Also, the actual boundary conditions at the well are replaced by a negative " line source " (sink) of constant intensity along the well-axis. The other boundary conditions are unaltered. Except in the immediate vicinity of the well and river, both the slope of the water-table and its lowering from the initial position as compared with the depth of the permeable medium will be small. The solution with the modified boundary conditions stated will therefore give an approximation to the actual conditions, except at points near the river and well, where the form of free-surface remains undetermined †. Moreover, the pressure distribution along the base of the permeable medium, or the average pressure and potential head over the vertical thickness of it, is probably obtained with sufficient accuracy for all practical purposes.

The general equations which are derived are unsuitable for ordinary practical use, as the integrals in them must be calculated by numerical integration. They have therefore been evaluated for a particular case to illustrate the most interesting characteristics of the flow—viz., the form of the water-table profile in a vertical plane through the well perpendicular to the river.

---

[*] See Muskat, 'Flow of Homogeneous Fluids,' McGraw Hill (1937), chap. ii. and iii., for a full discussion of the fundamental assumptions involved.

[†] From the present view-point this is not a serious limitation, as this paper is mainly concerned with the form of free surface at some distance from the well and river.

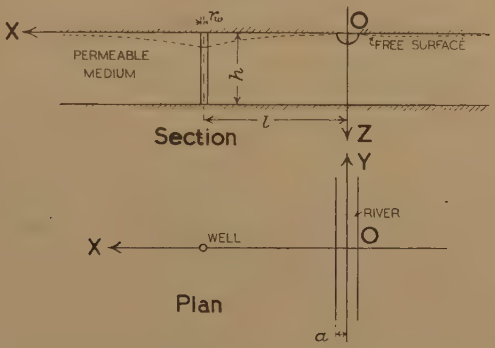
The practically important question as to the degree to which the river forms a "barrier" in restricting the lowering of the water-table on the side of the river away from the well is also considered.

The equations lead to a simple asymptotic formula from which the practical extent of the cone of depression can easily be calculated for any specified values of the dimensions and constants involved.

Mathematical Solution.

Let co-ordinate axes be chosen as shown in fig. 1. OY coincides with the axis of the half-cylinder of radius a representing the river bed, the positive direction of **Z** being vertically downwards. The axis of the pumped well is located at $x=l$, $y=0$. Let the permeable medium be bounded above and below by the horizontal impervious planes $z=0$ and $z=h$, respectively.

Fig. 1.



It will be convenient to write

$$\xi=x/h, \quad \eta=y/h, \quad \zeta=z/h, \quad \lambda=l/h, \quad \alpha=a/h,$$

so that ξ, η, ζ are co-ordinates measured in a unit equal to the vertical thickness of the permeable medium ; λ is the distance between the axes of the river and well, and α is the radius of the river-bed measured in the same unit.

A potential function ϕ is introduced, which is defined by

$$\phi=k\left(\frac{p}{g\gamma}-z\right), \qquad \text{. . . . .} \qquad (1)$$

where p is the pressure intensity in the pore water (atmospheric pressure being taken as zero), γ is the density of the water, and k is the "effective permeability" of the permeable medium *.

---

* Muskat, *loc. cit.* pp. 72, 258.

Darcy's law is expressed by the relations

$$v_x = -\frac{1}{h} \cdot \frac{\partial \phi}{\partial \xi}, \quad v_y = -\frac{1}{h} \cdot \frac{\partial \phi}{\partial \eta}, \quad v_z = -\frac{1}{h} \cdot \frac{\partial \phi}{\partial \zeta}, \quad . \quad . \quad . \quad (2)$$

v_x, v_y and v_z being the component velocities in the directions of x, y, and z, respectively.

The function ϕ must satisfy the following conditions :—

At all points of the permeable medium

$$\frac{\partial^2 \phi}{\partial \xi^2} + \frac{\partial^2 \phi}{\partial \eta^2} + \frac{\partial^2 \phi}{\partial \zeta^2} = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

At the impervious boundaries $\zeta = 0$ and $\zeta = 1$

$$\partial \phi / \partial \zeta = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

On the semi-cylindrical surface $\xi^2 + \zeta^2 = \alpha^2$

$$\phi = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

On the axis of the well $\xi = \lambda$, $\eta = 0$.

A sink of constant flux density * . . . . (6)

As $\xi \longrightarrow \infty$, or $\eta \longrightarrow \infty$, $\phi \longrightarrow 0$. . . . . (7)

To obtain a solution for ϕ, it is first supposed that the bounding planes $\zeta = 0$ and $\zeta = 1$ are removed, and that the medium extends to infinity in all directions.

Consider the potential function

$$\bar{\phi} = \text{A} \log \bar{\rho} + \text{B}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where $\bar{\rho}^2 = (\lambda - \xi)^2 + \eta^2$, $\bar{\rho}$ being the distance (measured in units of h) of any point from the axis of the well. Equation (8) satisfies (3) and gives the required line sink at the well. Since $\bar{\phi}$ is independent of ζ, (4) is also satisfied.

Let q be the volume of water flowing to unit length of the well in unit time, and let r_w be the radius of the well ; then

$$q = \frac{2\pi r_w}{h} \cdot \frac{d\bar{\phi}}{d\rho}_{\bar{\rho} = r_w/h} = 2\pi \text{A}, \text{ from (8).}$$

Therefore $\text{A} = q/2\pi$. Choosing the constant B so that $\bar{\phi} = 0$ at the origin

$$\bar{\phi} = \frac{q}{2\pi} \log \left(\frac{\rho}{\lambda}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

---

* The true boundary conditions at the surface of the well are : $\phi = -kz$, $0 \leqslant z \leqslant z_w$; and $\phi = -kz_w$, $z_w \leqslant z \leqslant h$; where z_w is the value of z at the pumping level in the well.

By assumption (*e*) (page 35), α and α/λ are small as compared with unity. Then, for any given value of η, the variation of $\bar{\phi}$ over the semi-circular boundary of the river-bed is small and, to a close approximation, the value of $\bar{\phi}$ at any point of the boundary equals the value on the axis. Thus

$$\bar{\phi}=\frac{q}{2\pi}\cdot\log\sqrt{\left\{1+\left(\frac{\eta}{\lambda}\right)^2\right\}},\quad \text{when } \xi^2+\zeta^2=\alpha^2. \quad . \; . \quad (10)$$

To satisfy condition (5) another potential function is now constructed, satisfying conditions (3) and (4) and giving values of ϕ on the river boundary cancelling those given by (10). It is first necessary to express (10) as a definite integral. It is known that

$$\log\sqrt{(a^2+x^2)}=\int_0^\infty\frac{1}{u}\,(e^{-u}-e^{-au}\cos xu)du\quad (a>0), \quad . \; . \; . \quad (11)$$

whence (10) becomes

$$\bar{\phi}=\frac{q}{2\pi}\int_0^\infty\frac{e^{-\lambda u}}{u}(1-\cos\eta u)du\quad \text{when } \xi^2+\zeta^2=\alpha^2. \quad . \; . \; . \quad (10a)$$

Consider the potential function

$$\phi_0'=\mathrm{K}_0(\rho_0 u)\cos\eta u, \quad . \; . \; . \; . \; . \; . \; . \; . \quad (12)$$

where K_0 denotes the modified Bessel function of zero order, and ρ_0 is the distance of any point (ξ, η, ζ) from the η-axis.

Equation (12) satisfies (3) when expressed in cylindrical co-ordinates, and since it represents a symmetrical distribution about the η-axis, it also satisfies condition (4) for the plane $\zeta=0$. To fulfil the condition for no flow across the plane $\zeta=1$, introduce a potential distribution symmetrical about the horizontal axis $\xi=0$, $\zeta=2$

$$\phi_2'=\mathrm{K}_0(\rho_2 u)\cos\eta u, \quad . \; . \; . \; . \; . \; . \; . \; . \quad (13)$$

where ρ_2 is the distance from this axis to the point considered (fig. 2). Equation (13) satisfies (3), while (12) and (13) taken together give a symmetrical potential distribution with respect to the plane $\zeta=1$, thus satisfying (4) at this plane.

Similarly, to cancel the flow across the plane $\zeta=0$, due to (13), requires

$$\phi'_{-2}=\mathrm{K}_0(\rho_{-2}u)\cos\eta u, \quad . \; . \; . \; . \; . \; . \; . \quad (14)$$

where ρ_{-2} is the distance from the axis $\xi=0$, $\zeta=-2$ to the point considered. Continuing in this way, and then adding (12), (13), (14), etc., it follows that

$$\phi'=\sum_{n=-\infty}^{\infty}\phi_{2n}=\sum_{n=-\infty}^{\infty}\mathrm{K}_0(\rho_{2n}u)\cos\eta u\quad (n=0 \text{ included}),$$

where $\rho_{2n}=\sqrt{\{\xi^2+(2n-\zeta)^2\}}$. Writing for shortness

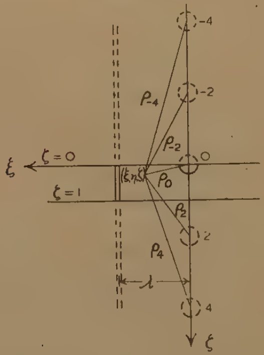
$$\mathrm{S}(\xi,\,\zeta)=\sum_{n=-\infty}^{\infty}\mathrm{K}_0(\rho_{2n}u).$$

Then $$\phi'=\mathrm{S}(\xi,\,\zeta)\,.\,\cos\eta u. \quad\ldots\ldots\ldots \quad (15)$$

Equation (15) satisfies (3) and (4), but it remains to consider the boundary condition at the river.

The predominant term in $\mathrm{S}(\xi,\,\zeta)$, for which $n=0$, gives a constant value of ϕ' over any semi-circular perimeter $\rho_0=\alpha$ of the river-bed. Also, since α/ρ_2 and α/ρ_{-2} are by hypothesis small, and the succeeding ratios rapidly diminish as n increases, the remaining terms will give a negligible variation in ϕ' over the perimeter $\rho_0=\alpha$. Hence the value of $\mathrm{S}(\xi,\,\zeta)$ at any point on the cylindrical boundary may, to a good approximation, be taken as $\mathrm{S}(0,\,\alpha)$. Then

Fig. 2.



$$\phi'=\mathrm{S}(0,\,\alpha)\cos\eta u,\quad\text{when }\xi^2+\zeta^2=\alpha^2.\quad\ldots\quad(16)$$

In order to cancel the potential distribution on the river-bed surface, represented by (10a), consider a more general solution formed from (15) and (16); viz.,

$$\phi'=-\frac{q}{2\pi}\int_0^{\infty}\frac{e^{-\lambda u}}{u}\left\{1-\frac{\mathrm{S}(\xi,\,\zeta)}{\mathrm{S}(0,\,\alpha)}\cos\eta u\right\}du.\quad\ldots\ldots\quad(17)$$

Assuming that the integral converges *, (17) satisfies (3) and (4) (the term in the integrand, independent of the co-ordinates not affecting $\partial\phi/\partial\zeta$). Moreover, when $\rho_0=\alpha$, it follows from (16) that ϕ' in (17) cancels $\bar{\phi}$ in (10 a). Hence the sum $\phi=\bar{\phi}+\phi'$ satisfies (5) as well as the other

---

* See p. 43.

boundary conditions. Then, by (9) and (17), the required solution is therefore

$$\phi=\frac{q}{2\pi}\left[\log\left(\frac{\bar{\rho}}{\lambda}\right)-\int_0^\infty e^{-\lambda u}\left\{\frac{\mathrm{S}(0,\,\alpha)-\mathrm{S}(\xi,\,\zeta)\,.\,\cos\eta u}{u\,\mathrm{S}(0,\,\alpha)}\right\}du\right], \quad . \quad (18)$$

where
$$\mathrm{S}(\xi,\,\zeta)=\sum_{n=-\infty}^{\infty}\mathrm{K}_0[u\sqrt{\{\xi^2+(2n-\zeta)^2\}}]$$

and
$$\bar{\rho}=\sqrt{\{(\lambda-\xi)^2+\eta^2\}} \qquad (n=0 \text{ included}).$$

Alternative Form of Solution.—While the integral in (18) can be evaluated without much difficulty for small values of ξ, an alternative form of solution is desirable when ξ is large. Another expression for the series $\mathrm{S}(\xi,\,\zeta)$ will therefore be obtained. Since

$$\int_0^\infty\frac{t\,.\,\mathrm{J}_0(at)dt}{k^2+t^2}=\mathrm{K}_0(ak)\,^*, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

it follows that

$$\mathrm{S}(\xi,\,\zeta)=\int_0^\infty t\mathrm{J}_0(t)\sum_{n=-\infty}^{\infty}\frac{1}{u^2\{\xi^2+(2n-\zeta)^2\}+t^2}\,.\,dt. \quad . \quad . \quad . \quad (20)$$

The original series in (18) converges for $u>0$; also, the uniform convergence of the integrand on the right-hand side of (20) can easily be established, and hence the change in the order of summation and integration justified.

Now the expression under the summation sign in (20) may be written in the form

$$\frac{1}{\sqrt{(u^2\xi^2+t^2)}}\,.\,\mathrm{R}\left\{\frac{1}{\beta-2nui}\right\},$$

where
$$\beta=\sqrt{(u^2\xi^2+t^2)}+iu\zeta.$$

Hence, the series in the integrand is equal to

$$\frac{1}{\sqrt{(u^2\xi^2+t^2)}}\,\mathrm{R}\left\{\frac{1}{\beta}+\sum_{n=1}^{\infty}\frac{2\beta}{\beta^2+(2nu)^2}\right\}, \quad . \quad . \quad . \quad . \quad (21)$$

which, by a well-known result †, equals

$$\frac{\pi}{2u\sqrt{(u^2\xi^2+t^2)}}\,.\,\mathrm{R}\left\{\coth\left(\frac{\pi\beta}{2u}\right)\right\}.$$

Thus (20) becomes

$$\mathrm{S}(\xi,\,\zeta)=\frac{\pi}{2u}\,.\,\mathrm{R}\int_0^\infty\coth\left(\frac{\pi\beta}{2u}\right)\frac{t\mathrm{J}_0(t)dt}{\sqrt{(u^2\xi^2+t^2)}}. \quad . \quad . \quad . \quad . \quad (22)$$

---

[*] Watson, ‘Theory of Bessel Functions’ (Camb. Univ. Press, 1922), p. 425, equation (5).

[†] Bromwich, ‘Infinite Series,’ 2nd ed. (Macmillan, 1926), p. 296.

Now
$$\coth z=(1+e^{-2z})(1+e^{-2z}+e^{-4z}+\ldots+e^{-2nz}+\ldots)$$
$$=1+2\sum_{n=1}^{\infty}e^{-2nz},$$

provided $|\arg z|<\pi/2$ to secure convergence. Hence, expanding the hyperbolic cotangent in the integrand of (22), and writing

$$\mathrm{R}(e^{-n\pi\beta/u})=e^{-\frac{n\pi}{u}\sqrt{(u^2\xi^2+t^2)}}\,.\cos n\pi\zeta,$$

it is found that

$$\mathrm{S}(\xi,\zeta)=\frac{\pi}{2u}\int_0^{\infty}\frac{t\mathrm{J}_0(t)}{\sqrt{(u^2\xi^2+t^2)}}\left\{1+2\sum_{n=1}^{\infty}e^{-\frac{n\pi}{u}\sqrt{(u^2\xi^2+t^2)}}\,.\cos n\pi\zeta\right\}dt. \quad . \quad (23)$$

It is known that *

$$\int_0^{\infty}t\,.\,\mathrm{J}_0(t)\frac{e^{-a\sqrt{(y^2+t^2)}}}{\sqrt{(y^2+t^2)}}dt=\frac{e^{-|y|\sqrt{(1+a^2)}}}{\sqrt{(1+a^2)}}. \quad . \quad . \quad . \quad (24)$$

(a positive).

Hence, evaluating the integrals in (23), it is found that

$$\mathrm{S}(\xi,\zeta)=\frac{\pi}{2u}e^{-u|\xi|}+\pi\sum_{n=1}^{\infty}\frac{e^{-|\xi|\sqrt{(u^2+n^2\pi^2)}}}{\sqrt{(u^2+n^2\pi^2)}}\cos n\pi\zeta. \quad . \quad . \quad (25)$$

The interchange of the order of summation and integration is justified, since the series is uniformly convergent with respect to t in the range of integration.

For values of $|\xi|$ greater than 2, the series (25) is rapidly convergent for all values of u.

Using (11) the log-term in (18) may be included in the integral. Thus

$$\log\left(\frac{\rho}{\lambda}\right)=\int_0^{\infty}\frac{e^{-\lambda u}}{u}\left\{1-e^{u\xi}\cos\eta u\right\}du \quad (\xi<\lambda). \quad . \quad . \quad . \quad (26)$$

Hence (18) may be written in the form

$$\phi=-\frac{q}{2\pi}\int_0^{\infty}\left\{\frac{\mathrm{S}(0,\alpha)-\mathrm{S}(\xi,\zeta)\,.\,e^{-\xi u}}{u\mathrm{S}(0,\alpha)}\right\}e^{-(\lambda-\xi)u}.\cos\eta u\,.\,du \quad . \quad . \quad (27)$$

when $\xi<\lambda$.

This formula is very convenient for fairly large negative values of ξ. Obtaining $\mathrm{S}(\xi,\zeta)$ from (25), all the terms except the first are then negligible, and

$$\mathrm{S}(\xi,\zeta)\,.\,e^{|\xi|u}\sim\pi/2u.$$

The values of ϕ obtained from (27) when ξ is negative can be used to

---

* Watson, *loc. cit.* p. 416, equation (2), from which (24) is easily deduced.

provide corresponding values when ξ is *positive*; for, it follows from (18) that

$$\phi(\xi,\ \eta,\ \zeta)=\phi(-\xi,\ \eta,\ \zeta)-\frac{q}{2\pi}\log\sqrt{\frac{(\xi+\lambda)^2+\eta^2}{(\xi-\lambda)^2+\eta^2}}. \quad . \quad . \quad . \quad . \quad (28)$$

In obtaining the two series for $S(\xi,\ \zeta)$, given in (18) and (25), an interesting and unusual relation is established between them.

Convergence of the Series and Integrals.—Since $K_0(x)$ tends exponentially to zero as x tends to infinity, the series in (18) for $S(\xi,\ \zeta)$ is convergent when $u>0$ for all values of ξ and ζ within the permeable medium. Hence the integrand in (18) is always finite for $u>0$. The factor $e^{-\lambda u}$ therefore secures convergence of the integral at the upper limit.

Now consider the behaviour of the integrand at the lower limit. Referring to (25), the contributions to both $S(\xi,\ \zeta)$ and $S(0,\ \alpha)$ of the terms under the summation sign are finite when $u\to0$. The contribution of the first term to $S(0,\ \alpha)-S(\xi,\ \zeta)\ .\ \cos\eta u$ is

$$\lim_{u\to0}\left[\frac{\pi}{2u}\left\{1-e^{-u|\xi|}\ .\ \cos\eta u\right\}\right]=\frac{\pi}{2}\ |\ \xi\ |,$$

which is finite if ξ is finite.

Also, $uS(0,\ \alpha)\to\pi/2$ when $u\to0$.

Hence $$\lim_{u\to0}\left\{\frac{S(0,\ \alpha)-S(\xi,\ \zeta)\cos\eta u}{uS(0,\ \alpha)}\right\}\text{ is finite.}$$

It follows that the integrand in (18) is finite at the lower limit. The convergence of the integral is thus established for finite values of ξ.

Asymptotic Formula for ϕ.—When ξ is negatively large, values of ϕ can easily be obtained from an asymptotic series. The detailed derivation of this series is rather long, so that only an outline of the proof and the final result will now be given. From (22) it can be shown that

$$uS(0,\ \alpha)=\frac{\pi}{2}-u\log\pi\alpha-\frac{\sigma\theta}{2\pi^2}\ .\ u^3, \quad . \quad . \quad . \quad . \quad . \quad (29)$$

where $$\sigma=\sum_{n=1}^{\infty}1/n^3\text{ and }0\leqslant\theta\leqslant1.$$

The first two terms in (29) give a good approximation to $S(0,\ \alpha)$ when u is small. Referring now to (27), it is clear that when ξ is negatively large the important part of the integral comes from the range near $u=0$. Hence, after substituting from (29) into the numerator and denominator of the integrand of (27), expanding the part in { } in powers of u, and evaluating the resulting simple integrals, it is found that

$$\phi_{\eta=0}=-\frac{q}{2\pi}\left\{\frac{c}{\xi'}-\left(\frac{c}{\xi'}\right)^2+\epsilon\right\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

where $$c=-\frac{2}{\pi}\log\pi\alpha, \qquad \xi'=\lambda-\xi$$

and $$|\epsilon|<\frac{2}{\xi'^3}\left\{|c^3|+\frac{\sigma}{\pi^3}\left(1-\frac{3c}{\xi'}+\frac{12c^2}{\xi'^2}\right)\right\}.$$

The corresponding values of ϕ for large positive values of ξ may then be found from (28).

It is clear from (30) that as ξ tends negatively to infinity, ϕ tends to zero when $\eta=0$. It then follows from (27) and (28) that ϕ tends to zero as $|\xi|$ or $|\eta|$ tend to infinity, thus satisfying boundary condition (7).

Approximation to the Form of the Free Surface.—At any point (x_0, y_0, z_0) of the free surface, the pressure p is zero; and hence by equation (1)

$$\phi(x_0, y_0, z_0)=-kz_0. \qquad (31)$$

In practice, z_0 will usually be small as compared with h, except in the immediate vicinity of the well. Also, recalling that $\partial\phi/\partial z=0$ when $z=0$, in the ideal problem which has been solved, it follows that

$$\phi(x_0, y_0, z_0)=\phi(x_0, y_0, 0) \text{ approximately.}$$

Substituting into (31), the surface $p=0$ is approximately given by

$$z_0=-\frac{1}{k}\phi(x_0, y_0, 0), \qquad (32)$$

where ϕ in (32) is obtained by putting $\zeta=0$ in (18) or (27).

Besides being a surface of constant pressure, the free surface is also a stream-surface, if seepage to it or evaporation from it are neglected *. When both z_0 and the slope of the free surface are small, this condition is approximately satisfied by (32), which may therefore be expected to give an approximation to the free surface except quite near to the well and river.

Numerical Evaluation of z_0.

The free-surface curve in the plane XZ is of special practical interest and has been determined for particular values of λ and α. This curve is found by putting $\eta=\zeta=0$ in (18) or (27) and substituting the resulting values of ϕ into (32) to give values of z_0.

In computing $S(0, \alpha)$ no appreciable error is introduced if α is put equal to zero in all terms after the first. Then

$$S(0, \alpha)\sim S_0=K_0(u\alpha)+2\sum_{n=1}^{\infty}K_0(2nu). \qquad (33)$$

---

* *Cf.* condition (*g*), p. 35.

The method of calculating $S(0, \alpha)-S(\xi, 0)$ in (18) depends on the magnitude of ξ. For sufficiently small values, it is convenient to expand $K_0(2nu)-K_0\{u\sqrt{(4n^2+\xi^2)}\}$ in powers of ξ, by Taylor's theorem, retaining only the first two terms. Then, using a well-known recurrence formula, it is found that

$$S_0-S(\xi, 0)\sim K_0(u\alpha)-K_0(u\xi) + \frac{\xi^2u^2}{2}\sum_{n=1}^{\infty}\{K_2(2nu)-K_0(2nu)\}. \quad . \quad . \quad . \quad . \quad (34)$$

TABLE I.

($\lambda=2$, $\alpha=0{\cdot}02$, $\eta=0$.)

$-\xi$.	$z_0 2\pi k/q$.	$z_m 2\pi k/q$.	$+\xi$.	$z_0 2\pi k/q$.	$z_m 2\pi k/q$.
0·00	—	0·545	0·00	—	—
·02	0·000	—	·02	0·000	—
·04	·118	—	·04	·158	—
·1	·270	—	·1	·370	—
·2	·362	·510	·2	·562	0·711
·4	·415	·480	·4	·821	·885
·8	·407	·429	·8	1·254	1·276
1·2	·384	·388	1·2	1·770	1·775
1·6	·353	·355	1·6	2·551	2·552
2·0	·327	·327	2±·002	(7·93)*	(7·93)
2·8	·283	—	2·8	2·075	—
3·6	·250	—	3·6	1·502	—
4·4	·224	—	4·4	1·205	—
5·2	·204	—	5·2	1·014	—
6	·186	—	6	·879	—
8	·153	—	8	·664	—
10	·131	—	10	·536	—
12	·114	—	12	·450	—
16	·090	—	16	·341	—
20	·075	—	20	·275	—

The series on the right-hand side of (33) and (34) are quickly evaluated, for simple values of u, from published tables of $K_0(x)$ and $K_2(x)$ †.

When ξ is too large for equation (34) to give a good approximation, equation (25) is used, two or three terms, at most, usually giving sufficient accuracy.

The integrals were evaluated for $\lambda=2$, $\alpha=0{\cdot}02$. Equation (34) was used for $\xi \leqslant 0{\cdot}4$, and (25) for $1{\cdot}2 \leqslant \xi \leqslant 4$. Values of the integrand in (18) were thus calculated for values of u, differing by 0·1 between 0 and 1·2 and by 0·2 between 1·2 and 3·6. The integral was then evaluated by Weddle's quadrature formula. It is easy to verify that neglect of the range $3\cdot6<u<\infty$ cannot cause an error in $z_0\times 2\pi k/q$ greater than 0·0001. For $\xi>4$, ϕ was likewise computed by quadratures from equation (27).

---

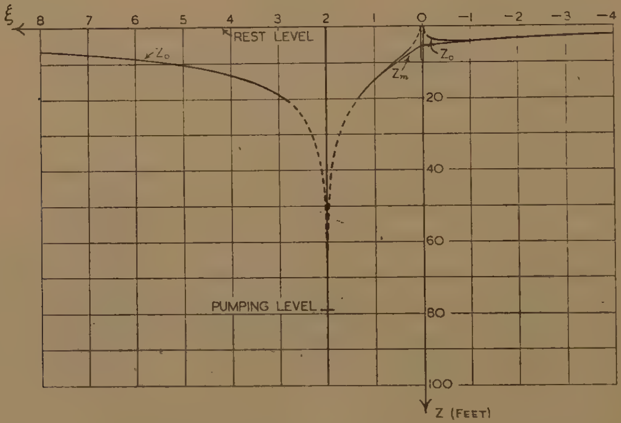
* This value is for $r_w=0{\cdot}002\,h$.

† Watson, *loc. cit.*

Values of the lowering z_0 of the water-table for $\eta=0$ are shown in Table I.

The values of $z_m=-\phi_m/k$, obtained by averaging ϕ over the thickness h of the permeable medium, are also given. (When $|\xi|>2$ these are practically equal to z_0). On evaluating $\int_0^1 S(\xi, \zeta)d\zeta$ from equation (25), the terms under the sigma sign vanish; and thus the average value of $S(\xi, \zeta)$ is given by the first term. Values of ϕ_m were then found from (27) by quadratures.

Fig. 3.



The results in Table I. depend only on the assumed values of α and λ. As an example of the effect of pumping from a well in (say) Bunter Sandstone, assume $k=3\cdot2\times10^{-5}$ ft./sec., $h=500$ ft., $l=1000$ ft., $a=10$ ft., $r_w=1$ ft., rate of pumping, $qh=1$ cu. ft./sec. Then $q/2\pi k=10$. The curves shown in fig. 3 were plotted from values obtained by multiplying the entries in Table I. by this factor. The z_0-curve is shown by a broken line near the well where z_0 ceases to be small compared with h, and hence the conditions for a good approximation are absent.

It will be observed that there is an appreciable lowering of the water-table as near to the river as $0\cdot04h$, where $z_0=1\cdot58$ ft. Considering the lowering on the side of the river remote from the well, the maximum value of z_0 is about 4·2 ft., and it occurs approximately at $\xi=-0\cdot5$.

For a given distance from the well, the value of z_0 is considerably greater on the side of the well away from the river than on the opposite side. For example, at a distance $4h$ from the well, $z_0=8{\cdot}79$ ft., in the direction away from the river, and 3·27 ft. in the opposite direction.

Results obtained from the asymptotic formula (30) may be compared with those from quadratures. Thus, by (30), when $\xi=-8$, $z_0=1{\cdot}451$ ft.; and when $\xi=-20$, $z_0=0{\cdot}737$ ft., the errors in z_0 being respectively less in absolute magnitude than 0·110 and 0·010 ft. The corresponding values by quadrature are 1·528 and 0·746, differing from the asymptotic values by less than the limiting errors just stated. This may be regarded as a check on equation (30).

Comparison with Two-Dimensional Solution.—The log-term in (28) represents the two-dimensional analogue of the three-dimensional problem considered in this paper. The former problem, which is readily solved by the "method of images" *, assumes a vertical bounding plane through the η-axis, in place of the river, over which ϕ is maintained constant. With the other boundary conditions as stated in (4), (6) and (7), the flow is then entirely horizontal. This two-dimensional solution will not in general give a good approximation to ϕ in the three-dimensional problem; moreover, it gives no indication of an important feature of the flow—the effect of pumping on the water-table beyond the river (ξ negative).

The values of z_0, for the two-dimensional solution, may easily be compared with those in Table I., since by (28) it follows that z_0, when ξ is negative, represents the difference between the two- and three-dimensional solutions for a point at the same distance $|\xi|$ in the positive direction.

The Value of α in a Practical Case.—Owing to the existence of a free-surface in the problem as originally stated, which has been neglected in the analysis, the solution obtained most nearly applies to a river-bed of cross-section somewhat wider than twice the depth (the semi-circular section). From considerations which are too detailed to be described here, it appears that for the numerical values adopted (p. 46), the width of river-bed is about 2·5 times the depth. Such a section is more probable in practice than the semi-circular section.

It is unnecessary to know the value of α very exactly, as z_0 is not very sensitive to variations in α. For large negative values of ξ, this follows immediately from formula (30). The first being the important term

$$|\phi|_{\eta=0} \propto \log \pi\alpha, \quad \text{approximately.}$$

Thus ϕ varies logarithmically with α when ξ is large.

---

* Muskat, *loc. cit.* p. 175, sec. 4. 7.

With the previous numerical values, the effect of doubling α is to decrease z_0 by approximately $-100\log_e 2/\log_e \pi\alpha$ per cent.; which for $\alpha=0{\cdot}02$ equals 25 per cent.

Allowance for Partial Puddling of the River-bed:—It has so far been assumed that the rock forming the river-bed has the same permeability k as the water-bearing formation. Actual river-beds are often "puddled" by the deposition of silt. If the river-bed is thus made practically impervious, the river will take no part in conditioning the flow of ground-water and the *steady* flow considered in this paper will not occur. If, however, the "puddling" is only partial, its effect may be approximately allowed for by using an equivalent value of α in equations (18) and (27).

Since ϕ is constant over the surface of the river-bed, the flow at this surface is entirely radial. If the puddled layer is not thick, it is reasonable to assume, as an approximation, that the flow is radial throughout this layer.

Let a be the actual radius of the river-bed, and a_1 the radius to the outside of the puddled part, so that the thickness of puddle $=a_1-a$.

Suppose now that the puddled layer, of effective permeability k_p, is replaced by rock of permeability k, allowance for the puddle being made by reducing the radius of the bed from a to a_e. Then a_e is determined by the condition that the decrease in total head (pressure+potential) between $\rho_0=a_e$ and $\rho_0=a_1$ with rock k shall equal the corresponding decrease between $\rho_0=a$ and $\rho_0=a_1$ with rock k_p; or, from equation (1), the decrease in ϕ/k must be the same in each case. Thus, if f is the rate of flow from the river per unit length (a function of η),

$$\frac{f}{\pi k}(\log a_e-\log a_1)=\frac{f}{\pi k_p}(\log a-\log a_1)\ *,$$

whence

$$\log (a_e/a_1)=(k/k_p)\log (a/a_1),$$

and the "equivalent radius"

$$a_e=a_1\ (a/a_1)^{k/k_p}. \qquad (35)$$

Hence, if k_p can be estimated, a_e is known. The equivalent value of α to be used then equals a_e/h.

The Extent of the "Cone of Depression."—From the asymptotic formula (30), a result of some practical importance can be deduced.

The extent of the cone of depression is best defined as the boundary outside which the lowering z_0 of the water-table is always less than some small assigned value (*e. g.*, 0·5 ft.).

Thus, confining attention to the vertical plane $\eta=0$, let ξ_1' be the distance (in h-units) from the well in the negative direction of ξ to the

---

[*] *Cf.* equation (9).

point where $z_0=\Delta$, and let ξ_2' be the corresponding distance in the positive direction.

Then, by (30),

$$\Delta \sim \frac{q}{2\pi k} \cdot \frac{c}{\xi_1'},$$

and ∴

$$\xi_1' \sim \frac{q}{2\pi k} \cdot \frac{c}{\Delta}, \quad \ldots \ldots \ldots \quad (36)$$

while by (28) and (30)

$$\Delta \sim \frac{q}{2\pi k}\left\{\frac{c}{\xi_2'+2\lambda}+\log\left(1+\frac{2\lambda}{\xi_2'}\right)\right\}.$$

Expanding in powers of the small quantity $2\lambda/\xi_2'$, and neglecting terms containing squares and higher powers,

$$\Delta \sim \frac{q}{2\pi k}\left(\frac{c}{\xi_2'}+\frac{2\lambda}{\xi_2'}\right)$$

or

$$\xi_2' \sim \frac{q}{2\pi k}(c+2\lambda)\frac{1}{\Delta}. \quad \ldots \ldots \ldots \quad (37)$$

TABLE II.

Values of ξ_2'/ξ_1'.

	$\lambda=1$.	$\lambda=2$.	$\lambda=4$.
$\alpha=0{\cdot}01$	1·91	2·82	4·63
$\alpha=\ {\cdot}02$	2·14	3·27	5·54
$\alpha=\ {\cdot}04$	2·51	4·03	7·06
$\alpha=\ {\cdot}06$	2·88	4·76	8·53

Dividing (37) by (36), and inserting the value of c,

$$\xi_2'/\xi_1' \sim 1-\lambda\pi/\log \pi\alpha. \quad \ldots \ldots \ldots \quad (38)$$

A few values of ξ_2'/ξ_1' are given in Table II.

The significance of the table is that, taking for example $\alpha=0{\cdot}02$, $\lambda=2$ (values on which Table I. is based), the "cone of depression" extends 3·27 times as far from the well in the direction away from the river as it does in the opposite direction. Assuming $\Delta=0{\cdot}5$ ft. and $q/2\pi k=10$, as previously, $\xi_1'=35{\cdot}2$ and $\xi_2'=115{\cdot}2$ (in h-units). These values show the extent of the "cone," as defined, in the directions $\mp x$.

Suppose now that the river-bed in this example is partially puddled for a thickness of 1 ft., having an effective permeability equal to one-tenth that of the permeable rock; then, if $h=500$ ft., it follows that $a=10$ ft., $a_1=11$ ft., $k/k_p=10$; and by equation (35), $a_e=4{\cdot}24$ ft. Finally, by (36) and (38), $\xi_1'=46{\cdot}2$, $\xi_2'=126{\cdot}2$. Hence, the assumed puddling does not greatly affect the " cone " in this case.

Summary and Conclusions.

(1) General equations for the motion of ground-water to a pumped well, under the conditions stated on pages 34–37, have been derived. (Equations (18), (27) and (28).) Numerical values for the lowering of the water-table have been calculated from these equations for a particular case. (Table I. and fig. 3.)

(2) A simple formula is given for the lowering of the water-table due to pumping, at large distances from the well. (Equation (30).)

The extension of the " cone of depression " in a direction away from the river relative to the extension in the opposite direction is expressed by a simple equation, and some typical values are stated. (Table II.)

(3) Other conditions remaining the same, an increase in the distance separating the well from the river increases the extension of the cone of depression in the direction away from the river, but it has no effect in extending the cone in the opposite direction beyond the river.

V. *The Birth and Descent of John Mayow: A Tercentenary Note.*

By Douglas McKie, D.Sc., Ph.D., Lecturer in the History of Science, University College, London *.

[Received November 17, 1941.]

Great doubt has for long existed among historians of science as to the date of John Mayow's birth, and dates ranging from 1640 to 1645, with a preference for the latter, are widely quoted; his birth-place has generally, and, as will presently be seen, quite wrongly, been given as London; and his ancestry has been referred to in most vague and uncertain terms. It is no part of the present writer's intention to record here all the statements that have been made on these matters, but rather to present evidence from sources that do not appear to have hitherto been taken into account and to present it in the order in which it has been encountered.

Anthony à Wood stated in the first edition of his 'Athenæ Oxonienses' that Mayow was "descended from a gentile family of his name living at *Bree* in *Cornwall*, was born in the parish of S. *Dunstan in the West* in *Fleetstreet*, *London*, admitted Scholar of *Wadham* Coll. 27. *Sept.* 1661 aged 16 years . . . ." (1). This statement, with the spelling of "gentile" amended to "genteel" and with some minor changes in the setting of the type, was repeated in the second edition of that notable compilation (2). While it has unfortunately given rise to the erroneous inference that Mayow was born in 1645 (since it states that he was aged 16 in 1661), and contains the apparently authoritative assertion that he was born in London, it happily provides us with a clue to the truth of these matters, though not a very obvious one.

If we turn to the work of C. S. Gilbert, we find that Mayow is described as "a descendant from the ancient and genteel family of his name, living at Bray, in the parish of Morval" (3). Mayow's territorial origins are thus made precise, and we have "a local habitation and a name" to guide us. The same author states later, when dealing with the family of Mayow of Bray, that: "From original documents, still preserved by this family, we find that Philip Mayow, merchant, at Looe, in the reign of Elizabeth, was the descendant of a younger branch of the Mayhews, (as the name was first written) of Devonshire, and supposed to have been nearly related to Jeremy Mayhew, of Boringdon, esq. the last of the elder line . . . . Philip Mayow, before mentioned of Looe, purchased in the sixth of Elizabeth, anno 1564, the manor of Bree, or Bray, in the parish of Morval, of Christopher Copplestone, of Warleigh, esq. which has since continued to be the principal residence of his descendants. He died in 1590, and was buried under the east end of the north aisle of St. Martyn's church, where an altar-tomb, bearing his full-length effigy, is still preserved" (4).

---

* Communicated by the Author.

It thus appears that Anthony à Wood's "Bree" is the manor of Bray in the parish of Morval near Looe in Cornwall. There John Mayow's family lived; and there they were succeeded by the Wynell Mayows, when the descent passed by the female line in 1766 to Philip Wynell (1716–81), who adopted his mother's maiden name of Mayow as an additional surname on succeeding his uncle Thomas Mayow (1682–1766). Philip Wynell was the son of John Wynell and Ursula Mayow, daughter of Phillip Mayow of Polgover (1640–1710), niece of John Mayow and sister of Thomas Mayow; John Wynell was the son of Philip Wynell, Vicar of Landrake, and Frances Mayow, sister of John Mayow. For these and other similar points, see the genealogical table given on p. 58. Some account of the manor of Bray is given by Thomas Bond (5), the historian of the Looes, and by Davies Gilbert, P.R.S. (6). Bond says: "In Cornish, Bray, Bre, Brea, signifies the Hill" (7); and this is evidently the same word as the Scots "brae."

The Phillip Mayow, who has been referred to above and who acquired the manor of Bray in 1564, died in 1590 and is buried in the Church of St. Martin's-by-Looe. His altar-tomb, on which he is represented in recumbent effigy wearing his alderman's robes, bears the epitaph: "Here lieth the bodye of Phillipe Maiowe of East Looe, Gentleman, who deceased this lyfe the 24th day of August in the yeare 1590 being then of 72 years" (8). A mural tablet, bearing the shield and crest of the Mayows, is flanked by the following interesting and curious verses:

> Here under this great carved stone
> Is Phillippe Maiow entombde,
> Who in his life for merchandice
> Was through this lande renownde.
>
> His trade was great, his dealins just,
> The poor did feel his bountie,
> Great cost he put for sea and land,
> In buildyng verye plentie.

With slight differences these verses are quoted by Bond (9) and Davies Gilbert (10), and more recently by Trevaldwyn (11). At one time Looe Island was included among the property of the Mayows of Bray (12). Polgover, near Bray, was one of their family seats, but later became a farm-house (13), as it is today, standing at a short distance from the manor-house of Bray, itself beautifully sited on the slopes of the lovely Bin Down or Bindon Hill. Other interesting details will be found in Trevaldwyn's book on St. Martin's Church, where we meet, in company with Phillip Mayow, many names famous in the West, Langdon and Buller, and even Thomas Bond himself, the historian of the Looes, who appointed Davies Gilbert, P.R.S., his residuary legatee and left him, stored in a cupboard in his house in Fore Street, East Looe, a sum of ten thousand guineas, which Gilbert carried in a cart to Liskeard (14). Further references to the Mayows of East Looe will be found in a recent study of the records of that borough (15). We might note, too, that we are here in the land of Richard Lower, John Mayow's great contemporary and friend; that a branch of the Wadham family lived not far away, one of them being several times Mayor of Liskeard (16); and that, also in John Mayow's day, the Lord Bishop of Exeter, of which diocese Cornwall was then part, was Seth Ward, F.R.S. Finally, with regard to the Phillip Mayow we have already mentioned, we might add that he was one of the nine Charter Burgesses when East Looe received its Charter of Incorporation from Queen Elizabeth on June 14, 1587 (17).

Turning now to the well-known work of Vivian (18), we learn that this Phillip Mayow of Looe, who purchased the manor of Bray in 1564 and died in 1590, *aet.* 72, was the son of another Phillip Mayow of Looe and was succeeded by his grandson Phillip (*ob.* 1659), having been predeceased by his son John, who had married Joan Rashleigh of Fowey (19), daughter of a house subsequently distinguished in the history of mineralogy in Cornwall. The grandson Phillip, who died in 1659, married Edith Vivian of Trelowarren, bearer of one of the most famous of Cornish names; he died without issue and was succeeded in Bray by his nephew Phillip (*ob.* 1677), a son of his younger brother John. Here at this point we have taken care to follow the advice of Vivian and Drake (20), warning their readers against assuming the accuracy of their transcripts: ". . . . we recommend the reader who descries a pearl of value, to have it verified by the clergyman of the parish . . . .". This advice would certainly seem to be necessary; for Vivian curiously gives the date of John Mayow's birth as 1645, although he had consulted the registers, and he must have missed this "pearl of value." The date given for the death of Phillip Mayow, 1677, has actually been corrected from 1679 after consulting the parish registers. Moreover, Vivian gives the date of the death of Phillip Mayow (*ob.* 1659) as 1658, ignoring the fact that a will dated March 7, 1658, O.S., is now to be reckoned as 1659. The Phillip Mayow of Bray, who succeeded his uncle in 1659 and died in 1677, married in Morval Church Frances Stuckley, daughter of John Stuckley of Afton and Frances, his wife, who was daughter and co-heiress of Sir John Leger and his wife, Katherine Neville, daughter of George Neville, Lord Abergavenny, himself descended from Edward I. The parish registers of Morval are fortunately extant; and the author is glad of this opportunity to express his deep appreciation of the kindness of the Rev. E. A. Saunders, M.A., Vicar of Morval, in searching these registers on his behalf, and also of his courtesy in affording him the privilege of verifying the relevant entries. First among these entries is the marriage of Phillip Mayow and Frances Stuckley, the parents of John Mayow. Under the entry for the year 1636, which Vivian gives incorrectly as 1635, it is recorded: "Nov. 22. Phillip Mayow gent. and Frances Stucley the daughter of John Stucley of Afton Esqr. were married." There follow the entries of the baptisms of their children. On October 29, 1637, a daughter, Edith, was baptized; and on January 22, 1639, which is Old Style and now to be reckoned as 1640, although Vivian gives it in his book without comment as 1639, their first son, Phillip, was baptized. Then there appeared their third child and second son, John, whom we take to be, with adequate reason, as will presently appear, the John Mayow, who subsequently became LL.D. and F.R.S., the chemist and physiologist, friend of Hooke and Lower. Under the entry for the month of December, 1641, the register reads: "John ye sonne of Mr Phillip Mayowe Gent. and Fraunces his wife was baptized ye 21 day."

Thus John Mayow was born in 1641 in the parish of Morval in Cornwall, and, unless we are to assume with some extravagance that he was born outside his parents' home, either in the manor-house of Bray, or, as his father had not yet succeeded his great-uncle Phillip in Bray, more probably in the neighbouring family seat, Polgover, by which name his elder brother was later known as Phillip Mayow of Polgover: and he was probably born in the month of December, since he was baptized on the 21st day of that month and since this ceremony was performed in those days as soon as possible after birth. It has been not a small part of the pleasure of carrying out these troublesome minutiae of historical research that the author has been able to establish, in what his work has also shown to be the tercentenary year, the Cornish birth of one of Cornubia's distinguished sons, so long and so groundlessly claimed for London.

Satisfactory as this solution may appear to be, the problem cannot, unfortunately, be left in this state. We have to recognize that we do not know today on what grounds the Mayow family based the information that they gave to Colonel Vivian in the eighties of last century identifying John Mayow, the son of Phillip Mayow and Frances Stuckley, with the Oxford chemist and physiologist. Some documents that might throw light on this matter and on Mayow's life in London are at the moment inaccessible and must remain so until the war is over. A most extraordinary contradictory statement occurs in the work of C. S. Gilbert, the very source of the clue that led the author to what he has set out here up to this point. Gilbert, in continuing his account of Phillip Mayow of Looe, who died in 1590, says: "He had issue two sons; Philip, who died in his father's life-time, but left issue; and John, who settled at Polruan, near Fowey, and dying at that place in 1645, was buried in Lantegloss church, where a marble monument remains to his memory. Philip, son of John, married in the year 1614, Edith, daughter of Hannibal Vyvyan, of Trelowarren, esq. and died without issue. In his will, dated 1658, he makes his kinsman, Philip, son of Philip Mayow, his sole executor, and residentiary legatee, which Philip, grandson of Philip, who died in 1590, and elder brother of John, and Stephen Mayow, married in 1636, Frances, daughter of John Stuckley . . . ."⁽²¹⁾.

There are a number of serious errors in this statement, and it carries a baffling foot-note, to which reference will be made later. Both text and foot-note unfortunately need detailed treatment in order to clear up a matter that has been greatly complicated by the extraordinary errors that occur in them. We shall first consider the passage quoted above. Reference to the genealogical table shows that the Phillip Mayow of Looe who died in 1590 was succeeded, not by his son, but by his grandson Phillip Mayow (*ob.* 1659); or, in other words, that Phillip Mayow of the second descent was succeeded by Phillip Mayow of the fourth descent. Further, it shows that John Mayow of the third descent was already dead in 1590, and certainly did not live until 1645; so that we may conclude that the John Mayow, whose monument is to be seen in Lanteglos Church, is another person altogether and not to be confused with John Mayow of the third descent of the Mayows of Bray; for otherwise he would have succeeded Phillip Mayow of the second descent as he was alive in 1590 on the latter's death. Continuing with Gilbert's statement, his "Philip, son of John" is also clearly not the son of the John Mayow of Polruan, who has been so unjustifiably dragged upon the scene: he is rather the son of the John Mayow of the third descent, and he died in 1659, as we have already shown. Finally, in the last sentence of the passage quoted certain changes are obviously needed and it should read: ". . . . his kinsman, Philip, son of John Mayow [fourth descent], his sole executor, and residentiary legatee, which Philip, great-grandson of Philip, who died in 1590, and son of John [fourth descent] and nephew of Stephen Mayow . . . .". After this analysis, we do not find ourselves anxious to extend any great confidence to any further statement by this author on the Mayows of Bray. However, the foot-note to which we have referred contains an assertion that we cannot ignore, even from an author whose credibility in this matter has thus been so seriously called into question.

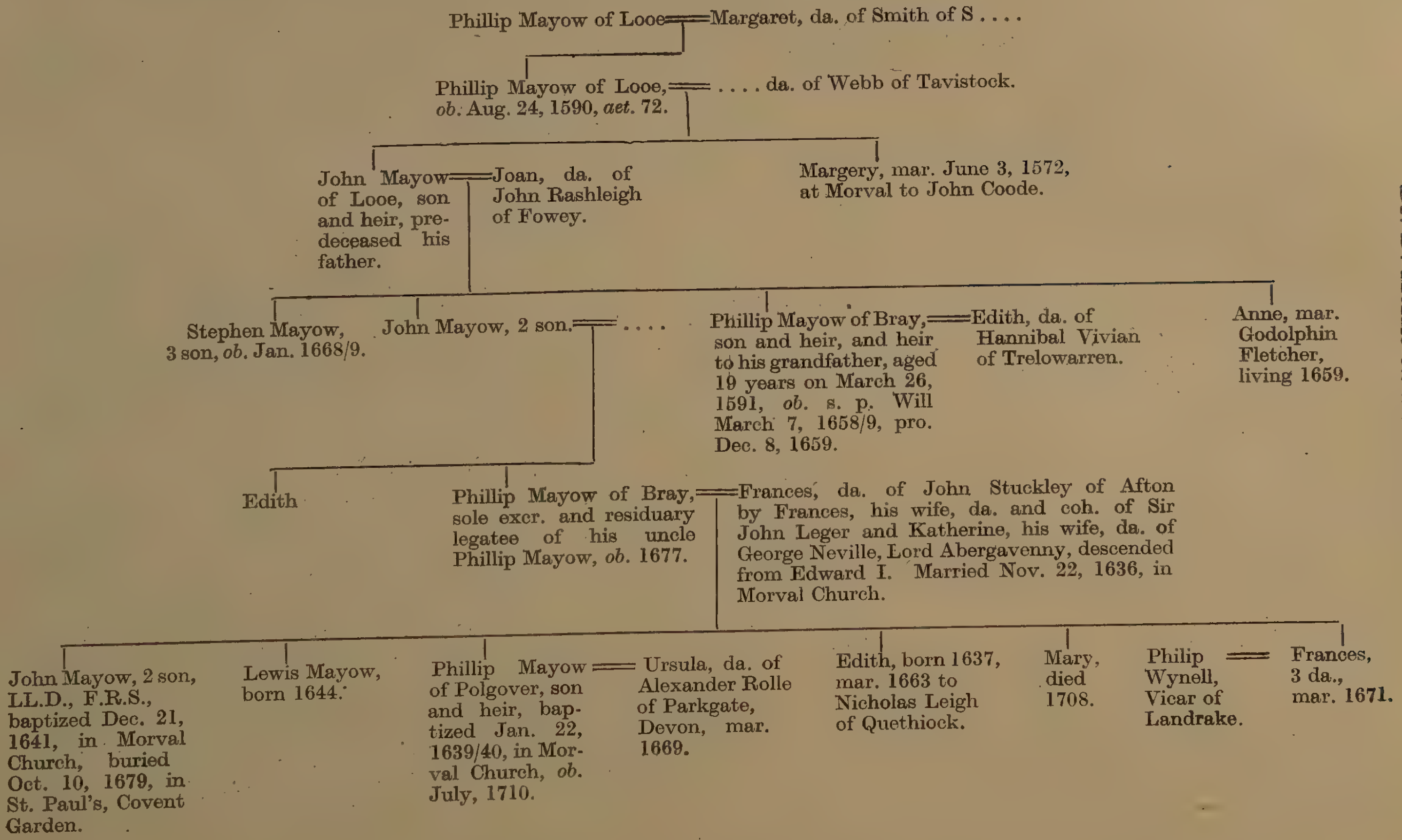
This extraordinary foot-note (22), referring to the John Mayow in the phrase "elder brother of John, and Stephen Mayow," reads: "This John, lived in London, at the time of the grand rebellion, and is supposed to have been the parliamentarian colonel Mayow, who is mentioned by Whitelock, as having been nearly murdered by the Portuguese ambassador's brother, and his party, in the streets, who mistook him for a Mr. Anstruther, who had incurred their displeasure. It is observed, that in this inhuman contest, which happened in 1653, colonel Mayow, gallantly defended himself, against this band of desperadoes; that he received seven dangerous wounds, and was left in a dying state. John, his son, born in London, in 1645, became a celebrated author, physician, and chemist, and has been already noticed in vol. I, page 140." The account of this attack on Colonel Mayow will readily be found in Whitelocke's 'Memorials' (23), where the name is given as Mayo. Cromwell's stout-hearted colonel is a historical figure, not to be charmed away; and it is not inherently obvious that Gilbert was guessing when he stated that the colonel was a Mayow of Bray and set down John Mayow, LL.D. and F.R.S., as his son.

In the face of this curious and entirely unsupported contradiction of almost all that these inquiries had so far established, it seemed not unreasonable to wonder whether Anthony à Wood and those who had followed him had actually exhausted, or indeed correctly reported, such information as might still be available in the extant records of Wadham College, Oxford, of which Mayow was a member. Happily, this suspicion proved correct. The author, through the good offices of his friend, Dr. K. J. Franklin, of Oriel College, has been provided by the Warden of Wadham College, to whom his grateful thanks are here expressed, with the following extract from the records of the College and with kind permission to quote it: "John Mayow (Mayouwe), matriculated 2 July 1658, received as commoner, but admitted scholar 23 Sept. 1659, said to be of Bree, Cornwall and aged 17." The John Mayow of Wadham College was therefore the John Mayow of Bray in Cornwall: and the John Mayow of Wadham College is also the John Mayow who, being aged seventeen in September of 1659 (he would have been eighteen in December), or possibly said to be of that age in July of 1658 (he would have been seventeen in December of that year), is, from every reasonable point of view, he who was baptized in Morval Church on December 21, 1641, the third child and second son of Phillip Mayow of Bray and Frances Stuckley, his wife. The John Mayow alleged to be born in London in 1645 in the Parish of St. Dunstan-in-the-West is a mere creature of Anthony à Wood, a ghost, who, it is hoped, is now successfully and permanently laid ; of him we may sincerely express the wish, *requiescat in pace,* for he has been a truly troublesome fellow.

It has, however, been a matter of considerable regret that all efforts to trace the details of Mayow's earlier education have so far failed. Inquiries have been made of many schools that existed in that period and whose records are extant, and other inquiries are still proceeding, but the publication of this work may draw the attention of those who have access to old school records and John Mayow's name may perhaps yet be found in one of such lists in the period from about 1650 to 1658, in which latter year he entered the University of Oxford. The author will be glad to hear of any scrutinies of any such lists, whether successful or not. The portrait of Mayow used for the engraving that appeared as the frontispiece of his 'Tractatus quinque,' published in 1674, has long been lost : it seems probable that it became the property of Mayow's wife, whom he married. according to Anthony à Wood (24), shortly before his death. Attempts to trace this marriage have so far not been successful : apart from the fact that the name of the bride is not known, the times are " out of joint " for work of this kind, which must be left for other days.

In concluding, we cannot, however, pass over one curious fact that has emerged in the course of this work. It has always been said that Hooke was numbered among Mayow's London friends, which is not to be wondered at, since they were much interested in the problems of respiration and combustion. Mayow died in London, although he was engaged in the practice of medicine in Bath during the summer ; and he was buried in the Church of St. Paul, Covent Garden, on October 10, 1679, as " John Mayo : Docter of Phisick. In the Church," according to the entry in the registers (25), although, in fact, he was LL.D., not M.D. Anthony à Wood tells us that he died "in an Apothecaries house, bearing the sign of the *Anker* in *Yorkstreet* near *Covent Garden,* within the liberty of *Westminster* . . . . in the month of *Sept.* in sixteen hundred seventy and nine, and was buried in the Church of S. *Paul* in *Covent Garden*" (26). Yet Robert Hooke, Mayow's friend, then resident in Gresham College and, so far as his almost daily entries in his diary show, moving daily among his associates and friends among the Fellows of the Royal Society, records for Saturday, November 8, 1679 : " Mayow Dead a month since" (27). Hooke records having been in company with Mayow on various dates at Childs and Garaways mostly ; the diary gives for these meetings the dates Oct. 29, Nov. 11, Dec. 4, 8, 10 and 15, in 1674, Jan. 11 and Feb. 11 in 1674/5, Dec. 13 in 1675, March 11 in 1675/6, and Nov. 1 in 1677. The nothing further appears, so far as the author has been able to trace, until the last entry of " Mayow Dead a month since." In the meantime, however, Hooke had proposed Mayow for the Fellowship

Genealogical table showing the descent of John Mayow, LL.D., F.R.S.; simplified and corrected from the table given by Vivian.



Phillip Mayow of Looe = Margaret, da. of Smith of S . . . .

Phillip Mayow of Looe, *ob.* Aug. 24, 1590, *aet.* 72. = ... da. of Webb of Tavistock.

John Mayow of Looe, son and heir, predeceased his father. = Joan, da. of John Rashleigh of Fowey.

Margery, mar. June 3, 1572, at Morval to John Coode.

Stephen Mayow, 3 son, *ob.* Jan. 1668/9.

John Mayow, 2 son. = ...

Phillip Mayow of Bray, son and heir, and heir to his grandfather, aged 19 years on March 26, 1591, *ob.* s. p. Will March 7, 1658/9, pro. Dec. 8, 1659. = Edith, da. of Hannibal Vivian of Trelowarren.

Anne, mar. Godolphin Fletcher, living 1659.

Edith

Phillip Mayow of Bray, sole excr. and residuary legatee of his uncle Phillip Mayow, *ob.* 1677. = Frances, da. of John Stuckley of Afton by Frances, his wife, da. and coh. of Sir John Leger and Katherine, his wife, da. of George Neville, Lord Abergavenny, descended from Edward I. Married Nov. 22, 1636, in Morval Church.

John Mayow, 2 son, LL.D., F.R.S., baptized Dec. 21, 1641, in Morval Church, buried Oct. 10, 1679, in St. Paul's, Covent Garden.

Lewis Mayow, born 1644.

Phillip Mayow of Polgover, son and heir, baptized Jan. 22, 1639/40, in Morval Church, *ob.* July, 1710. = Ursula, da. of Alexander Rolle of Parkgate, Devon, mar. 1669.

Edith, born 1637, mar. 1663 to Nicholas Leigh of Quethiock.

Mary, died 1708.

Philip Wynell, Vicar of Landrake. = Frances, 3 da., mar. 1671.

of the Royal Society on Feb. 7, 1677/8, and Mayow had been elected on Nov. 30, 1678 (28). For two years, therefore, Hooke, who filled his diary almost every day with ample entries, often with mere names, records no meeting with Mayow, who, however, died at no very great distance from Hooke's own dwelling. Moreover, Hooke's diary shows that he himself was in London and following his usual avocations during the period in question. This may suggest some estrangement between old friends closely associated in the study of the chemical problems of respiration and combustion, indeed between neighbours, and an estrangement so great that a month passed before Hooke knew that Mayow was dead. Such a difference would, on certain grounds, seem very remarkable. On this and on some other points it is hoped that documents not now accessible may be able to throw some light when the war is over.

In conclusion, the author wishes to express his thanks for assistance to Miss Wynell Mayow of Bray, and to the Rev. W. M. M. Picken, M.A., Rector of St. Martin's-by-Looe ; and also to his friend, C. C. James, Esq., of Penzance, for his help in searching the labyrinths of Cornish topographical literature.

References.

(1) 'Athenæ Oxonienses,' London, 2 vols., 1691–2 : ii, p. 474.

(2) London, 2 vols., 1721 : ii, p. 637.

(3) 'An Historical Survey of the County of Cornwall' etc., Plymouth-Dock and London, 2 vols. in 3, 1817–20 : i, p. 140.

(4) *Ibid.*, ii, p. 197.

(5) 'Topographical and Historical Sketches of the Boroughs of East and West Looe' etc., London, 1823, pp. 183–4.

(6) 'Parochial History of Cornwall' etc., London, 1838, 4 vols. : iii, p. 250. Davies Gilbert took his account of the Looes and the districts bordering upon them from Bond's book, which has been already referred to ; these two authors were distantly related, and Mrs. Davies Gilbert drew the sketches of East Looe and West Looe for Bond's book, on p. 308, of which they are indexed as "Embellishments."

(7) Bond, *op. cit.*, p. 183.

(8) B. W. J. Trevaldwyn, 'A Short Account of the Church and Parish of St. Martin-by-Looe,' Plymouth, 1901, p. 49. It might be noted, at this point, that the spelling of the name "Philip" varies considerably among the different documents and authors quoted here : these have been left as they are, without any attempt to introduce a dull uniformity.

(9) *Op. cit.*, p. 21.

(10) *Op. cit.*, iii, p. 250.

(11) *Op. cit.*, p. 49.

(12) Bond, *op. cit.*, p. 30 ; Davies Gilbert, *op. cit.*, iv, p. 37.

(13) Bond, *op. cit.*, p. 184 ; Davies Gilbert, *op. cit.*, iii, p. 252.

(14) Trevaldwyn, *op. cit.*, p. 55.

(15) A. L. Browne, 'Corporation Chronicles. Being Some Account of the Ancient Corporations of East Looe and of West Looe in the County of Cornwall,' Plymouth, 1904, pp. 18, 31, 36 (*bis*), 42, 50.

(16) J. Allen, 'History of the Borough of Liskeard and its Vicinity,' London, 1856, p. 476.
(17) Bond, *op. cit.*, pp. 2–3.
(18) J. L. Vivian, 'The Visitations of Cornwall, comprising The Heralds' Visitations of 1530, 1573 & 1620. With Additions by Lieutenant-Colonel J. L. Vivian,' Exeter, 1887, pp. 315–6, "Mayow of Bray": see also J. L. Vivian and H. H. Drake, "The Visitation of the County of Cornwall, in the Year 1620," 'The Publications of the Harleian Society,' London, 1874, ix, pp. 140–1.
(19) See also Vivian, *op. cit.*, p. 391 : the marriage took place on February 3, 1570, at Fowey.
(20) *Loc. cit.*, p viii.
(21) *Op. cit.*, ii, pp. 197–8.
(22) *Ibid.*, p. 198.
(23) B. Whitelocke, 'Memorials of the English Affairs' etc., London, 1682, p. 569 : new edn., Oxford, 1853, 4 vols., iv, pp. 49–50.
(24) *Op. cit.*, 1st edn., ii, p. 475 ; 2nd edn., ii, p. 637.
(25) 'The Publications of the Harleian Society, Registers,' 1908, xxxvi, p. 85. The Rector of St. Paul's, Covent Garden, the Rev. R. A. de V. Hart-Davies, has kindly informed the author that there is no stone or tablet marking the place of Mayow's burial.
(26) *Op. cit.*, 1st edn., ii, p. 475 ; 2nd edn., ii, p. 637.
(27) 'The Diary of Robert Hooke' etc., edited by Henry W. Robinson and Walter Adams, London, 1935, p. 430.
(28) T. Birch, 'The History of the Royal Society of London' etc., London, 4 vols., 1756–7 : iii, pp. 384, 442.

VI. *The Influence of Diffusion on the Propagation of Shock Waves.*

By T. G. Cowling, M.A., D.Phil. (Manchester) *.

[Received August 25, 1941.]

Abstract.

It is shown that in the propagation of shock waves through a gas-mixture diffusion produces effects similar to those of viscosity and thermal conduction. If the molecular weights of the two gases are not very different, the effects are small : when they are as widely different as those of hydrogen and oxygen, the diffusion effects are more important than those of thermal conduction, and are comparable with those of viscosity. The velocities of diffusion involved are an appreciable fraction of the velocity of sound in the gas.

(1) It is well known that the equation of energy in a shock wave can be satisfied only when dissipative forces are taken into account. The effects of thermal conduction and viscosity on the propagation of the wave have been considered by Rankine, Rayleigh, Taylor and others (see Lamb, 1932, p. 487). We shall show in this paper that diffusion in a binary mixture of gases has an effect similar to that of these other free-path phenomena, and that, if the gases are of very different molecular weights, its importance is comparable with theirs.

Diffusion arises in a shock wave because of differences of pressure : its direction is such that the heavier molecules tend to move toward the regions of greater pressure. Thus the lighter molecules move more slowly through the region of variable density than the heavier ones, and the gas in this region contains an increased proportion of lighter molecules. It is because of the effect of this change in composition on the equation of state that diffusion is able to influence the motion. The pressure diffusion is small if the masses m_1, m_2 of molecules of the two gases are nearly equal, or if the proportion of either gas in the mixture is small ; in such a case diffusion is not important in a shock wave.

(2) In order to isolate the effect in question, we ignore viscosity, thermal conduction, and thermal diffusion. We consider a plane wave

---

* Communicated by Professor S. Chapman, F.R.S.

advancing without change of type: let the densities of the gas in front of and behind the wave disturbance be uniform and equal to $1/v'$ and $1/v''$, where $v' > v''$. Take axes moving with the velocity of the wave-front, with Ox normal to the wave-front and pointing into the region of density $1/v''$; then the velocity u at any point is positive, and the density $1/v$ and pressure p increase steadily with x. Since the motion is steady, the mass crossing unit area of a plane x=const. in unit time is independent of x. Call it M; then

$$u = \mathrm{M}v, \quad \dots\dots\dots\dots \quad (1)$$

which is the equation of continuity.

There are also equations of continuity for the two gases separately. Let f, $1-f$ be the proportions by volume of the two gases in the mixture at any point. Then their proportions by mass are

$$m_1 f/\{m_1 f + m_2(1-f)\}, \quad m_2(1-f)/\{m_1 f + m_2(1-f)\}$$

or

$$(1+\mu)f/(\mu+f), \quad \mu(1-f)/(\mu+f),$$

where

$$\mu = m_2/(m_1 - m_2); \quad \dots\dots\dots\dots \quad (2)$$

we suppose, as we may, that $m_1 > m_2$, so that μ is positive. Thus if u_1, u_2 are the mean velocities of molecules of the two gases, the equations of continuity for the two gases are

$$\frac{fu_1}{\mu+f} = \frac{f_0 \mathrm{M}v}{\mu+f_0}, \quad \frac{(1-f)u_2}{\mu+f} = \frac{(1-f_0)\mathrm{M}v}{\mu+f_0}, \quad \dots\dots \quad (3)$$

where f_0 is the value of f in the regions of constant density.

The total momentum or energy transported across unit area of a plane x=const. in unit time is independent of x. The momentum is the sum of the molecular and the molar momenta; thus the equation of momentum is

$$p + \mathrm{M}^2 v = \text{const.} = \mathrm{A}, \text{ say.} \quad \dots\dots\dots\dots \quad (4)$$

The energy-transport is made up of (*a*) transport of mass kinetic energy $\frac{1}{2}\mathrm{M}u^2$; (*b*) transport of heat energy with the mass, of amount

$$pu\{f/(\gamma_1 - 1) + (1-f)/(\gamma_2 - 1)\},$$

where γ_1, γ_2 are the ratios of specific heats for the two gases; (*c*) transport of heat-energy through the mass, of amount q, say; and (*d*) work pu done by the pressure. Hence, adding these,

$$pu\left\{\frac{f\gamma_1}{\gamma_1 - 1} + \frac{(1-f)\gamma_2}{\gamma_2 - 1}\right\} + q + \tfrac{1}{2}\mathrm{M}u^2 = \text{const.} = \mathrm{B}, \text{ say.}$$

In this q denotes the flux of energy due to diffusion alone; it is given by

$$q=p\left\{\frac{f\gamma_1}{\gamma_1-1}(u_1-u)+\frac{(1-f)\gamma_2}{\gamma_2-1}(u_2-u)\right\}$$

(Chapman and Cowling, 1939, p. 145, equation (3): $\gamma_1/(\gamma_1-1)$ and $\gamma_2/(\gamma_2-1)$ appear instead of the factor 5/2 given there, because the equation there refers only to monatomic gases; the diffusion term involving the thermal diffusion coefficient is neglected). Substituting this value of q, the energy equation becomes

$$p\left\{\frac{f\gamma_1 u_1}{\gamma_1-1}+\frac{(1-f)\gamma_2 u_2}{\gamma_2-1}\right\}+\tfrac{1}{2}\mathrm{M}u^2=\mathrm{B}.$$

or, using (1) and (2),

$$p\mathrm{M}v\frac{\mu+f}{\mu+f_0}\left\{\frac{f_0\gamma_1}{\gamma_1-1}+\frac{(1-f_0)\gamma_2}{\gamma_2-1}\right\}+\tfrac{1}{2}\mathrm{M}^3v^2=\mathrm{B}. \quad . \quad . \quad . \quad (5)$$

The values of A and B in (4) and (5) are found by using the fact that $f=f_0$ when $v=v'$ or $v=v''$. Substituting them and simplifying, we get

$$p=\mathrm{M}^2\left\{\frac{\kappa}{1+\kappa}(v'+v'')-v\right\}; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

$$\frac{f_0-f}{f_0+\mu}=\frac{\kappa(v'-v)(v-v'')}{v\{\kappa(v'+v'')-v(1+\kappa)\}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where

$$\kappa=f_0\frac{\gamma_1+1}{\gamma_1-1}+(1-f_0)\frac{\gamma_2+1}{\gamma_2-1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The velocity of diffusion u_1-u_2 is given by

$$u_1-u_2=\frac{\mathrm{D}_{12}}{f(1-f)}\left\{\frac{f(1-f)}{(\mu+f)p}\frac{dp}{dx}-\frac{df}{dx}\right\},$$

where D_{12} is the coefficient of diffusion (Chapman and Cowling, 1939, p. 144, equation (7), and p. 140, equation (7)). Substituting for u_1 and u_2 from (3), this becomes

$$\mathrm{M}v(\mu+f)\frac{f_0-f}{\mu+f_0}=\mathrm{D}_{12}\left\{\frac{f(1-f)}{(\mu+f)p}\frac{dp}{dx}-\frac{df}{dx}\right\}. \quad . \quad . \quad . \quad (9)$$

Equations (6), (7), and (9) determine the space-variations of pressure, density, and composition of the gas in the wave.

(3) Since p is positive when $v=v'$, we have from (6)

$$v'/v''<\kappa. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

If $\gamma_1=\gamma_2=\gamma$ this becomes identical with Rayleigh's condition that v'/v'' cannot exceed $(\gamma+1)/(\gamma-1)$ derived for a shock wave in which viscosity is the only dissipative force.

There are, however, further restrictions on the possible values of v'/v''. The expression on the right of (7) is positive everywhere within the wave disturbance, and so $f<f_0$. Hence the proportion $(1-f)$ of lighter molecules is greater there than in the regions of uniform density. We must, however, have f positive everywhere. Now it can be shown that the maximum value ϕ of the right-hand side of (7) for values of v between v' and v'' is the smaller root of the equation

$$\left(\frac{v'-v''}{v'+v''}\right)^2=\frac{(\kappa-1)\phi-\kappa\phi^2}{\kappa-(\kappa+1)\phi}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and, to secure that f is always positive, we must have

$$\phi<f_0/(f_0+\mu). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

This condition in general provides a further limitation on the possible values of v'/v''. It can, however, be shown that ϕ is always less than $(\kappa-1)/(\kappa+1)$; thus, if f_0/μ is greater than $\frac{1}{2}(\kappa-1)$, (12) is automatically satisfied.

A more stringent upper limit to v'/v'' is obtained from (9). Since $f<f_0$ in the wave, we have

$$\frac{df}{dx}<\frac{f(1-f)}{(\mu+f)p}\frac{dp}{dx}.$$

Using (6) and (7), and remembering that dv/dx is negative, we can transform this into

$$\frac{v^2(v'+v'')-2(1+\kappa)vv'v''+\kappa v'v''(v'+v'')}{v^2\{\kappa(v'+v'')-(1+\kappa)v\}}<\frac{(1+\kappa)f(1-f)}{\kappa(\mu+f)(\mu+f_0)}. \quad . \quad (13)$$

The expression on the left is positive near $v=v''$, but negative near $v=v'$, and decreases steadily as v increases from v'' to v'; that on the right is everywhere positive, and is small only if f is small. Hence, if the inequality is violated, this must occur at or near $v=v''$, unless f/f_0 is small for some value of v between v'' and v'. Putting $v=v''$ in (13), we get

$$\frac{v'-v''}{v''}<\frac{(1+\kappa)f_0(1-f_0)}{\kappa(\mu+f_0)^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

This condition is very restrictive if the molecular masses m_1, m_2 are nearly equal, since then μ is large. For example, in the atmosphere, regarded as a mixture of four parts nitrogen and one part oxygen, it indicates that v'/v'' cannot exceed 1·0037. Thus diffusion is ineffective as a controlling agency in shock waves in the atmosphere, and the same is generally true if m_1/m_2 is not very different from unity. Since f does not vary greatly with v in this case (the maximum variation in f is roughly proportional to $(m_1-m_2)^3$) the use of (14) in place of (13) is justified.

If m_1/m_2 is large, v'/v'' can take much larger values, and greater variations in f are possible. In this case it is no longer certain that (14) is the least favourable case of (13), and further study of (13) is necessary. When v'/v'' is large, the left-hand side of (13) falls off rapidly with increasing v both near $v=v''$ and near $v=v'$; near its zero between these values it decreases relatively slowly. The minimum value of the right-hand side (if one exists) occurs at the zero of the left-hand side *. Hence, roughly, (13) is violated either if (14) is violated or if the minimum value of the right-hand side is very small; the second alternative implies that f/f_0 is very small. If (12) is satisfied with a certain margin of safety (so that f/f_0 is nowhere extremely small) and (14) is satisfied, we may expect that (13) will also be satisfied.

For example, in the following table we give the maximum values of v'/v'' for which (12), (14), and (13) are satisfied for mixtures of hydrogen

f''.	v'/v'' for (12).	v'/v'' for (14).	v'/v'' for (13).
0·8	6	1·25	1·25
0·6	6	1·63	1·63
0·4	6	2·28	2·28
0·2	6	3·62	3·62
0·133	5·8	4·37	4·37
0·1	5·25	4·78	4·78
0·067	4·43	5·08	4·37
0·033	3·19	4·76	3·15
0·0083	1·82	2·71	1·80
0·00067	1·2	1·17	1·17
0·0001	1·07	1·03	1·03

and oxygen of different proportions. It is clear from this table that (13) follows very closely the more stringent of (12) and (14); it is in general identical with (14) when this is the more stringent, and slightly more restrictive than (12) when this is the more stringent.

The figures given in the table indicate that large changes in density in shock waves are permissible as a result of diffusion when the molecular masses are very unequal and the proportion of each gas in the mixture is not small. Diffusion is more effective than thermal conduction in such a case, since the latter does not permit v'/v'' to exceed $(3\gamma-1)/(\gamma+1)$, which for the above mixture is 4/3. Viscosity permits all values of v'/v'' up to $(\gamma+1)/(\gamma-1)$; diffusion does not permit such large values in the above case, but they would be possible for a greater mass-ratio m_1/m_2.

---

* The right-hand side of (13) has a minimum only if f decreases below $\sqrt{(\mu^2+\mu)}-\mu$, or $m_2^{\frac{1}{2}}/(m_1^{\frac{1}{2}}+m_2^{\frac{1}{2}})$: the minimum corresponds to $df/dv=0$, and the left-hand side of (13) equals $-\dfrac{p(1+\kappa)}{M^2(\mu+f_0)}\dfrac{df}{dv}$.

There is observational evidence of the existence of velocities comparable with the velocity of sound in stellar atmospheres ; in the resulting shock waves the mutual diffusion of ions and electrons would be important.

(4) Another measure of the relative efficiency of diffusion, thermal conduction and viscosity is the thickness of the layer in which the transition from the initial density to the final is made. Substituting for p and f from (6) and (7) into (9), we obtain an equation connecting dv/dx and v ; this can be integrated, remembering that D_{12} is proportional to v *. The integration, though involving only elementary integrals, is complicated ; it gives

$$\begin{aligned}-\frac{\kappa M x}{\kappa+1}=\frac{D_{12}}{v}\Bigg\{&\frac{\kappa}{\kappa+1}\log\frac{\kappa v'v''-v^2}{(v'-v)(v-v'')}\\&+(\mu+f_0)^{-2}\Bigg[\frac{f_0(1-f_0)}{v'-v''}\{v''\log(v-v'')-v'\log(v'-v)\}\\&+\tfrac{1}{2}\{\kappa(\mu-2f_0\mu-f_0)+\kappa^2\mu(\mu+1)\}\log(\kappa v'v''-v^2)\\&+\frac{\mu(1+\mu)(v'+v'')}{4(v'v'')^{\frac{1}{2}}}\kappa^{\frac{3}{2}}\log\frac{(\kappa v'v'')^{\frac{1}{2}}+v}{(\kappa v'v'')^{\frac{1}{2}}-v}\\&-\tfrac{1}{2}\mu(1+\mu)\kappa^2\frac{(v'+v'')v-(1+\kappa)v'v''}{\kappa v'v''-v^2}\Bigg]\Bigg\}+\text{const.}\quad . \ . \ . \quad (15)\end{aligned}$$

Thus, for example, for mixtures of oxygen and hydrogen, if $v'/v''=2$ and $f_0=\frac{1}{4}$, and the gas of density $1/v'$ is at N.T.P., we find that the transition from $v=\frac{1}{10}v''+\frac{9}{10}v'$ to $v=\frac{9}{10}v''+\frac{1}{10}v'$ occurs in a region of width $1{\cdot}23\times10^{-4}$ cm. The corresponding width when viscosity alone is considered is $3{\cdot}6\times10^{-5}$ cm. Thermal conduction alone does not permit so large a value of v'/v'' ; but if $v'/v''=1{\cdot}3$ the widths of the similar transition zone are $3{\cdot}8\times10^{-4}$ cm. considering diffusion alone, and 8×10^{-5} cm. considering thermal conduction alone. Thus in these examples diffusion permits a more gradual transition between the initial and final states than either viscosity or thermal conduction, and so is more effective than either of these. The comparison is, however, somewhat misleading in the case of viscosity ; for the changes in density may be more abrupt near $v=v''$ for diffusion than for viscosity, even though the reverse is true of the average over the whole range.

(5) The extreme nature of the conditions in a shock wave can perhaps be better illustrated by considering the diffusion effects than by con-

---

* Actually D_{12} varies with the composition slightly, and to a greater extent with the temperature. The neglect of all variation save that with the density is parallel to Rayleigh's assumption of constant thermal conductivity and viscosity.

sidering those of viscosity and thermal conduction. In the wave disturbance there is a considerable variation of composition; *e. g.*, in a mixture of hydrogen and oxygen with $f_0=\frac{1}{4}$ and $v'/v''=2$, the minimum value of f is about 0·2. By (8) and (3), this implies that the velocity of diffusion u_1-u_2 is an appreciable fraction of u, or, since u is the velocity of the wave-front relative to the gas, an appreciable fraction of the velocity of sound; in the above example u_1-u_2 is about $\frac{1}{4}u$ at the point corresponding to the minimum f, and this is about 0·3 of the velocity of sound in the region of density $1/v'$. Thus it is clear that the velocity of diffusion is not much smaller than the velocity of thermal agitation of the molecules. Thermal conduction and viscosity would produce a similar large effect on the distribution of the molecular velocities; but this is not so clearly apparent from an inspection of the numerical results.

References.

Lamb, 1932, 'Hydrodynamics,' Sixth Edition, Cambridge University Press.
Chapman and Cowling, 1939, 'The Mathematical Theory of Non-uniform Gases,' Cambridge University Press.

VII. *The Sedimentation of Washed Red Blood-Cells.*

By R. B. Whittington, M.Sc.*
(of the Engineering Department, The University, Manchester).

[Received October 1, 1941.]

Abstract.

The experiments to be described were made as a first step towards a mechanical investigation of the red-cell sedimentation test.

The sedimentation of whole-blood suspensions presents a complex mechanical picture which may be simplified by suspending the washed red-cells in an isotonic saline solution.

The cylindrical strings of cells known as rouleaux do not form when the plasma has been washed away; and we are left with discrete cells of known mean size and density falling through a liquid of known viscosity and specific gravity.

List of Symbols.

F=hydrodynamic resisting force or drag.
ψ=drag-coefficient.
a=projected area of a single cell.
v=velocity at any instant.
v_t=" terminal " or maximum velocity.
v_s=Stokes' velocity.
ρ=density of fluid.
ρ'=mean density of cells.
μ=coefficient of absolute viscosity of fluid.
u_s=coefficient of absolute " effective viscosity " of suspension.
ν=coefficient of kinematic viscosity of fluid=μ/ρ.
$d=2r$=mean projected diameter of a cell.
V=mean volume of a cell.
R_e=Reynolds' Number $\dfrac{vd}{\nu}$.
ϕ=volume of cells per unit volume of suspension.
M=mean mass of a cell.

$$\left.\begin{array}{l}\dot{x}=\text{velocity}\\ \ddot{x}=\text{acceleration}\end{array}\right\}\text{of fall.}$$

---

* Communicated by J. Allen, D.Sc.

THE "terminal" sedimentation-velocities must depend on the values of the drag-coefficient ψ defined by Newton's equation

$$F=\psi \, . \, a \, . \, v^2 \, . \, \rho.$$

Stokes' solution for rigid spheres in a fluid of infinite extent is

$$\psi=\frac{12}{R_e},$$

leading to the well-known result at the terminal velocity

$$v_s=\frac{2}{9}\frac{(\rho'-\rho)}{\mu} \, . \, r^2. \quad \text{(" Stokes' Law.")}$$

It seems most unlikely that red blood-cells, being elastic biconcave discs approximating to short cylinders of length roughly 1/6 of their diameter, should have the same drag-coefficients as rigid spheres.

In order to investigate the drag-coefficients of red-cells one must obviously observe the cells at various velocities. Varying the specific gravity of the fluid, or altering the "effective viscosity" of the suspension, are the methods which present themselves. However, the fluid must be kept isotonic *, so that, in the absence of a range of isotonic liquids of widely-varying specific gravities, the only possible method is to vary the "effective viscosity" of suspension by varying the proportion of suspended cells.

Hatschek's (1) relation

$$\mu_s=\frac{\mu}{1-\sqrt[3]{\phi}}$$

is only intended to hold good for values of ϕ greater than about 50 per cent.

Hess (2) found that

$$\mu_s=\frac{\mu}{1-a\phi},$$

where a is a numerical coefficient varying from about 2·25 at $\phi=10$ per cent. to about 1·15 at $\phi=79$ per cent.

Hess found these coefficients for suspensions of red blood-cells (see fig. 1); and his results agree closely with those calculated from Hatschek's

---

* *I. e.*, isotonic with blood-plasma. A sodium chloride solution, 0·85 per cent. by weight, fulfils this requirement.

If the cells are suspended in a solution containing an excess of salt they shrink owing to a loss of fluid by osmosis. If the dispersion fluid contains a deficiency of salt the cells take in fluid, increase in size, and if the deficiency of salt be great enough they burst.

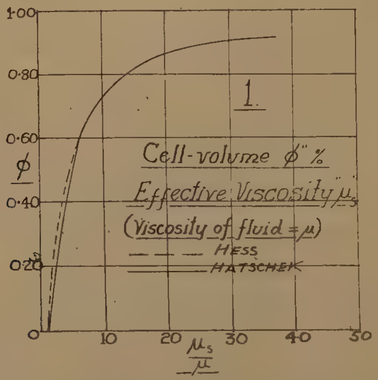
Thus a hypertonic medium causes shrinkage of the suspended cells, while a hypotonic one leads ultimately to hæmolysis, *i. e.*, disruption of the cells.

formula over the range ($\phi > 50$ per cent.) for which the latter applies. Hatschek's formula was tested with red blood-cells by Trevan (3), and close agreement between the theoretical and experimental results was obtained.

All the values of ϕ in the author's experiments are less than 50 per cent., it having been found that denser concentrations of washed cells sedimented too slowly to be observed. For when red blood-cells have been subjected to the washing treatment described below, they show a marked increase in fragility, and in high concentrations it was found that hæmolysis tended to take place before the sedimentation-velocities could be measured.

The "effective viscosities" have therefore been calculated from Hess's formula, and are to some extent conventional, as will be remarked upon later. Fig. 1 shows both Hess's and Hatschek's formulæ over the whole range of all concentrations from $\phi = 0$ to $\phi = 90$ per cent.

Fig. 1.



In the sedimentation-experiments now to be described, Westergren pipettes of approx. 2·52 mm. bore and with a capacity of 1 c.c. were used.

In each experiment a quantity of blood, either 5, 10, 15 or 20 c.c., was withdrawn from a vein and mixed in the syringe with 20 per cent. of a 3·8 per cent. solution of sodium-citrate as anti-coagulant. In the first experiment (I) a small quantity of venous blood was taken without citrate for a red-cell count.

The citrated blood was then centrifuged in conical tubes until the cells were packed. The plasma was pipetted away and replaced by normal saline. The suspensions were then well shaken with the object of breaking down rouleau-formation, and the centrifuging repeated. The supernatant liquid was again withdrawn and the washing process repeated once more.

Finally, the washed cells were mixed with saline in varying proportions, the volumes of cells per unit volume of suspension being determined by centrifuging in graduated tubes.

This was the method of volume-determination used by Trevan in verifying Hatschek's formula.

The various suspensions were then taken up into Westergren-tubes for sedimentation.

The specific gravity of the red cells was determined by weighing, the mean value being 1·09 gm./c.c.

In each case the dispersion fluid was 0·85 per cent. saline, whose kinematic viscosity-coefficient was determined in a capillary viscometer of a modified Ostwald type, which had been designed to determine the viscosities of small quantities (about 1·2 c.c.) of blood plasma, and had been previously calibrated against a series of pure liquids.

The viscosity of the saline was measured at several temperatures and its value at the mean sedimentation-temperature interpolated.

TABLE I.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ref. No.	ϕ.	$\frac{\mu_s}{\mu}$	μ_s.	v_t.	R_e.	ψ.
I/1	0·105 } $\mu = 1{\cdot}067 \times 10^{-2}$	1·31	$1{\cdot}40 \times 10^{-2}$	$7{\cdot}03 \times 10^{-5}$	$35{\cdot}7 \times 10^{-7}$	$3{\cdot}87 \times 10^{6}$
I/2	0·152	1·50	1·60 ,,	5·12 ,,	22·8 ,,	7·27 ,,
I/3	0·155	1·52	1·62 ,,	4·83 ,,	21·3 ,,	8·20 ,,
I/4	0·360	2·86	3·05 ,,	1·40 ,,	3·26 ,,	97·3 ,,
I/5	0·407	3·41	3·64 ,,	0·913 ,,	1·78 ,,	230 ,,
I/6	0·303	2·38	2·54 ,,	2·35 ,,	6·58 ,,	34·6 ,,
II/1	0·415 } $\mu = 1{\cdot}053 \times 10^{-2}$	3·52	3·71 ,,	0·967 ,,	1·86 ,,	204 ,,
II/2	0·445	4·00	4·21 ,,	0·585 ,,	0·986 ,,	557 ,,
II/3	0·083	1·24	1·31 ,,	7·55 ,,	41·0 ,,	3·36 ,,
III/1	0·119 } $\mu = 1{\cdot}074 \times 10^{-2}$	1·36	1·46 ,,	6·45 ,,	31·4 ,,	4·59 ,,
III/2	0·211	1·77	1·90 ,,	3·40 ,,	12·7 ,,	16·6 ,,
III/3	0·347	2·75	2·96 ,,	1·70 ,,	4·08 ,,	66·1 ,,
III/4	0·358 } $\mu = 1{\cdot}074 \times 10^{-2}$	2·83	3·04 ,,	1·46 ,,	3·39 ,,	89·6 ,,
IV/1	0·269	2·12	2·28 ,,	2·50 ,,	7·79 ,,	30·6 ,,

Notes.

Column (1).—The numerals I, II, III, IV denote the blood sample.

,, (2).—ϕ=cell concentration. With 100 per cent. cells $\phi = 1{\cdot}00$.

,, (3).—μ_s/μ=ratio of effective viscosity of suspension to viscosity of dispersion fluid. Hess's formula.

,, (4).—μ_s=coefficient of absolute effective viscosity of suspension. Gm./cm.sec.

,, (5).—v_t=maximum or " terminal " velocity of fall. Cm./sec.

,, (6).— R_e=Reynolds' Number. Dimensionless. Measured by $\left(\frac{v_t \cdot d \cdot \rho}{\mu_s}\right)$, where $\rho = 1{\cdot}00$ gm./c.c.

,, (7).—ψ=Drag Coefficient. Dimensionless. Measured by $\frac{\text{Volume } (\rho' - \rho) g}{\text{Projected area} \times v_t^2 \cdot \rho}$

The packed cell-volume of the whole blood was found by centrifuging to be 44·6 per cent. (after correcting for citrate), and the red-cell count on the same sample was found to be $5{\cdot}19\times10^6$ per cub. mm.

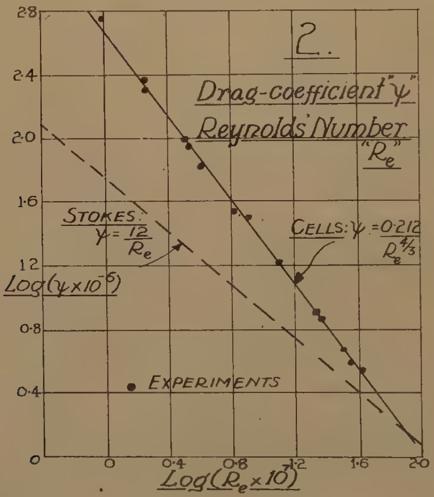
The mean volume of a single cell was therefore

$$\frac{0{\cdot}446}{5{\cdot}19\times10^6}\,\text{mm.}^3 \text{ or } 86\mu^3.$$

Mogensen (4) gives the average normal cell-volume as precisely 86 cubic microns, and the average normal cell-diameter in isotonic fluid as 7·11 μ.

It was therefore assumed that the cell-diameter in each experiment (conducted with blood from the same subject) averaged 7·11 μ, and that the mean volume of a single cell averaged 86 μ^3.

Fig. 2.



It may be pointed out here that the drag-coefficient results are again conventional in the sense that the volume of 86 μ^3 takes no account of the mode of packing in the graduated tubes.

From the results set out in Table I. the drag-coefficients were plotted logarithmically against the Reynolds' Numbers (fig. 2), and it was found that the experimental points were fairly represented by the line

$$\psi=\frac{0{\cdot}212}{R_e^{4/3}}.$$

Now putting our solution

$$\psi=\frac{0{\cdot}212}{R_e^{4/3}}$$

into Newton's equation $F=\psi\,.\,a\,.\,v^2\,.\,\rho$, and assuming that $a = \pi \frac{d^2}{4}$, we get

$$F = \frac{0{\cdot}212}{R_e^{4/3}} \pi \frac{d^2}{4} \cdot v^2 \cdot \rho, \quad \text{where } R_e = \frac{vd}{\mu_s}$$

$$= \frac{0{\cdot}053\pi(vd)^{2/3}\, \mu_s^{4/3}}{\rho^{1/3}}.$$

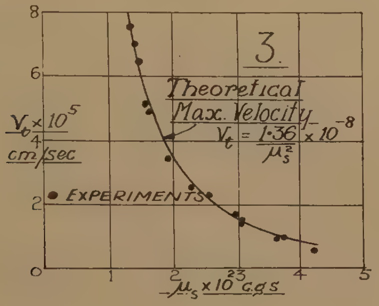
The net accelerating force on each cell is thus

$$V(\rho' - \rho)g - \frac{0{\cdot}053\pi(vd)^{2/3}\, \mu_s^{4/3}}{\rho^{1/3}},$$

and writing $V = 86 \times 10^{-12}$ c.c.; $(\rho' - \rho) = 0{\cdot}09$ gm./c.c.; $g = 981$ cm./sec.^{2}; $d = 7{\cdot}11 \times 10^{-4}$ cm.; $\rho = 1{\cdot}00$ gm./c.c.; we have

$$F = 7{\cdot}59 \times 10^{-9} - 13{\cdot}3 \times 10^{-4}(v \cdot \mu_s^2)^{2/3}. \quad . \quad . \quad . \quad . \qquad (1)$$

Fig. 3.



At the terminal velocity the acceleration is zero, so that

$$(\mu_s^2 v_t)^{2/3} = \frac{7{\cdot}59 \times 10^{-9}}{13{\cdot}3 \times 10^{-4}},$$

$$\text{or } v \doteq \frac{1{\cdot}36 \times 10^{-8}}{\mu_s^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

or, more generally,

$$v_t = \left[\frac{V(\rho' - \rho)g}{0{\cdot}166}\right]^{3/2} \times \frac{\rho^{1/2}}{d \cdot \mu_s^2}.$$

In fig. 3 the equation (2) is plotted out, and the experimental points from Table I. columns (4) and (5) are also shown.

We may obtain an approximate integration of equation (1) in any particular case.

In equation (1), writing $F=M\ddot{x}$ where the mass M of a single cell is equal to

$$1{\cdot}09\times 86\times 10^{-12}\ \text{gm., then}$$

$$\ddot{x}=\frac{7{\cdot}59\times 10^{-9}-13{\cdot}3\times 10^{-4}\,(\dot{x}\,.\,\mu_s{}^2)^{2/3}}{1{\cdot}09\times 86\times 10^{-12}}$$

$$=80{\cdot}9-14{\cdot}2\times 10^{6}\,(\mu_s)^{4/3}\,(\dot{x})^{2/3}. \qquad (3)$$

Example.—Ref. No. II/3. $\mu_s=1{\cdot}31\times 10^{-2}$ c.g.s.

$$\ddot{x}=80{\cdot}9-4{\cdot}38\times 10^{4}\,(\dot{x})^{2/3}. \qquad (4)$$

Here $\ddot{x}$ becomes zero when $\dot{x}=7{\cdot}92\times 10^{-5}$ cm./sec. (See Table II. column (2).) Approximate integration of equation (4) shows that 99 per cent. of this terminal velocity is reached in the very short time of $4{\cdot}31\times 10^{-6}$ secs.

TABLE II.

(1)	(2)	(3)	(4)	(5)	(6)
Ref. No.	Terminal Velocity " v_t." Theoretical.	Experimental.	v_s.	v_s/v. (Theoretical.)	v_s/v_t. (Experimental.)
I/1	$6{\cdot}93\times 10^{-5}$	$7{\cdot}03\times 10^{-5}$	$17{\cdot}7\times 10^{-5}$	2·55	2·52
I/2	5·31 ,,	5·12 ,,	15·4 ,,	2·90	3·01
I/3	5·18 ,,	4·83 ,,	15·3 ,,	2·94	3·18
I/4	1·47 ,,	1·40 ,,	8·08 ,,	5·49	5·77
I/5	1·02 ,,	0·913 ,,	6·76 ,,	6·59	7·42
I/6	2·12 ,,	2·35 ,,	9·70 ,,	4·59	4·14
II/1	0·986 ,,	0·967 ,,	6·66 ,,	6·72	6·89
II/2	0·765 ,,	0·585 ,,	5·86 ,,	7·65	10·0
II/3	7·92 ,,	7·55 ,,	18·9 ,,	2·38	2·50
III/1	6·38 ,,	6·45 ,,	16·9 ,,	2·65	2·62
III/2	3·77 ,,	3·40 ,,	13·0 ,,	3·44	3·82
III/3	1·55 ,,	1·70 ,,	8·34 ,,	5·36	4·90
III/4	1·48 ,,	1·46 ,,	8·13 ,,	5·49	5·57
IV/1	2·61 ,,	2·50 ,,	10·8 ,,	4·14	4·34

Notes.

Column (2).—Theoretical maximum velocities (cm./sec.) calculated from $v_t=\dfrac{1{\cdot}28}{\mu_s{}^2}\times 10^{-8}$.

,, (3).—Experimental maximum velocities (cm./sec.).

,, (4).—Terminal velocity for a rigid sphere of dia. 7·11 μ (Stokes) (cm./sec.).

,, (5).—v_s/v_t using the values in column (2) for v_t.

,, (6).—v_s/v_t ,, ,, (3) for v_t.

Until packing of the cells becomes appreciable, therefore, the theoretical distance/time curves are linear, the initial acceleration period being insensibly short *.

---

* Experimentally, the position of the upper boundary of the red-cell layer was observed at suitable intervals; the fall of this boundary was plotted against the time, and the slope of the linear portion of the curve was measured, giving the terminal velocity of fall.

Stokes' equation is, as we have seen,

$$v_s = \frac{2/9\, g(\rho' - \rho)r^2}{\mu_s},$$

and for spheres of 7·11 μ diameter and density 1·09 gm./c.c. we get

$$v_s = \frac{274{\cdot}2 \times 10^{-8}}{\mu_s} \text{ cm./sec.} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Values of the Stokes' velocity from equation (5) are set out in Table II. column (4).

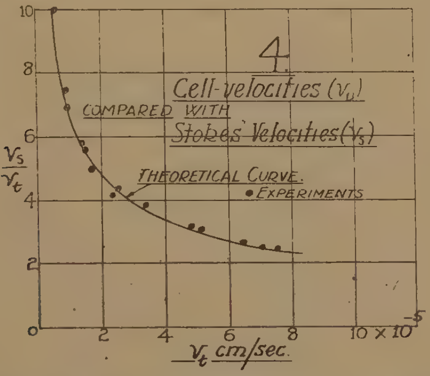
Suppose at a certain viscosity μ_s, $v_t = v_s$, that is, the red-cells fall with the velocity given by Stokes' law.

We have $\dfrac{1{\cdot}36 \times 10^{-8}}{\mu_s^2}$ (from equation (2))

must equal $\dfrac{247{\cdot}2 \times 10^{-8}}{\mu_s}$ (equation (5)),

or $\mu_s = 0{\cdot}55 \times 10^{-2}$ c.g.s. units.

Fig. 4.



Since the absolute viscosity of the saline is greater than 1×10^{-2} c.g.s. units, it is impossible to reach the Stokes' velocity except by using some isotonic liquid of very low viscosity *.

The fact that the Stokes' and the red-cell drag-coefficient lines intersect does not, of course, indicate that the cell-velocity should equal the Stokes' velocity at this point: since, for a given viscosity of suspension, the Stokes' R_e is higher and the drag-coefficient lower than the corresponding quantities for the red-cells.

However, the convergence of these lines indicates that the cell and Stokes' velocities approach one another, as shown in fig. 4. The cells

---

* The author does not know of any low-viscosity liquid which would not cause distortion of the cells.

become relatively more "streamlined" as the velocity increases; and this may be wholly or partly due to increasing elastic deformation of the cells. As the transverse velocity-gradient increases, the viscous shear increases, and the cells may be pulled out of shape into a more "streamlined" form.

Any tendency for the cells to deform elastically is no doubt implicit to some extent in the values of the Hess coefficients; that is, Hess's "μ_s" may be not merely a constant function of a given suspension, but may also be some function of the transverse velocity-gradients in his viscometer.

The values of the drag-coefficients given here are thus somewhat conventional. Apart from the boundary-effects, the diameter of the sedimentation-tube may play a part in determining the shape of the falling cells, by the effect of the tube-diameter on the transverse velocity-gradients.

Again, as has been pointed out, the mode of packing * of the centrifuged cells is unknown, so that the assumed volume is also somewhat arbitrary.

However, the drag-coefficients evaluated here enable the sedimentation-velocities of washed cells to be forecast, provided that the same methods of measurement are used.

They appear, moreover, to be suggestive of the line of approach to the study of whole-blood sedimentation; and such sedimentation is in considerable use as a clinical test, though the mechanics of the problem are not yet understood.

The author wishes to thank Professor Baker (of the Department of Pathology in the University of Manchester), Dr. Miller (Medical Superintendent of the Crossley Sanatorium), and Dr. Herring (Assistant-Director of the Clinical Pathology Laboratories, Manchester Royal Infirmary), for their kindness in providing laboratory facilities and invaluable technical help and advice; and others who have most kindly obtained the necessary blood-samples.

Bibliography.

(1) E. Hatschek, 'The Viscosity of Liquids' (1928). ('International Text-Books of Exact Science.')
(2) W. R. Hess, *Koll.-Zeitsch.* xxvii. p. 1 (1920).
(3) J. W. Trevan, Biochem. J. xii. p. 60 (1918).
(4) Erik Mogensen, 'The Size of the Red Blood-cells, etc.' (1938). (H. Milford.)

---

* The extreme elasticity of the red-cells suggests that they pack very much more closely than rigid bodies, and it seems probable that the volume of liquid occluded between the cells is relatively small.

VIII. *Notices respecting New Books.*

The Biologic Fundamentals of Radiation Therapy. By FRIEDRICH ELLINGER. Translated by REUBEN GROSS. [Pp. xvi+360.] (New York: Elsevier Publishing Co., 1941. London Distributors: H. K. Lewis & Co. Price 30*s*.)

THIS book is a translation, revised and brought up to date, of the author's original German edition published in 1935. Written for medical men and biologists, its interest for physicists is chiefly confined to those engaged on radiation work. The book gives an ordered account of a very extensive experimental field, in which physical as well as biological sciences are involved. It will be of great value for all who are concerned with the theory or practice of radiation therapy. The effect of every type of radiation, electromagnetic and corpuscular, on normal and pathological tissue is dealt with comprehensively in the first four sections of the book, and a fifth section is devoted to theories of the biological action of radiations and general principles of radiation therapy.

The physicist will be particularly grateful that so much biological information is made available in one book. He will, however, have some misgivings as to the presentation and interpretation of the physical basis of radiation dosage, and on this aspect of the book some comment should be made. Leaving radiation physics to other text-books, the author summarizes the properties of the various radiations in three pages, and thereafter includes definitions of physical dose units as footnotes. The unit of X-ray dose, the roentgen, is thus defined correctly in terms of X-ray absorption in air; the remarks on p. 15, however, carry the completely erroneous implication that the roentgen measures the incident rather than the absorbed X-ray energy—a rather common and unfortunate misunderstanding of the significance of the unit. There are other obscurities, as on p. 276, in connexion with the Schwarzchild photochemical equation, and on p. 284, where a paragraph on exposure to ultraviolet rays does not make clear what relation, if any, the author implies between the Schwarzchild time factor and the inverse square law.

The book is well printed, is illustrated with excellent plates and diagrams, and includes an extensive bibliography of 1100 references covering the period since 1930. It will find deserved popularity among the readers to whom it is primarily addressed. The comments of the reviewer on the treatment of physical aspects of radiation therapy have been made to underline the importance of a proper understanding of physical dose measurement, and the necessity for that close co-operation between the radio-therapist, biologist and physicist which Dr. Ellinger himself advocates. F. W. S.

The Nature of the Chemical Bond. By LINUS PAULING. Second Edition. [Pp. xvi+450.] (Ithaca, New York: Cornell University Press. London: Oxford University Press, 1940. Price 28*s*.)

THERE can be no better indication of the value of this book than the fact that a second edition has been found necessary so soon after the appearance of the first edition, the preface of which bears the date June 1938. No radical changes have been made in the contents, but two new sections dealing with restricted rotation about single bonds and the conditions for equivalence or non-equivalence of bonds have been added, and a number of additional references are given to investigations carried out since 1938.

The book will continue to be of interest and value to a wide circle of readers. Pauling himself has an equal claim to be considered a physicist, a chemist or a crystallographer, and it is certain that physicists, chemists and crystallographers opening Pauling's book with quite different objects in view will all find a great deal that is of first rate interest to them. But in addition to these more specialized readers, others who are less advanced, undergraduates in chemistry and perhaps chemists brought up on more classical lines, will find the book a source of inspiration and a guide in this newly developed field of quantum chemistry ; the clarity and simplicity of the treatment are such that the less advanced reader will not be discouraged at the start by mathematical complexities, and in fact it might almost be said that the theoretician is more likely to be repelled by unfamiliar chemical formulæ than the experimentalist, physicist or chemist, by mathematical formulæ.

It remains only to express a warm welcome for the second edition of this valuable work. G. W. B.

A Short History of Science to the Nineteenth Century. By Charles Singer. [Pp. xiv+399.] (Oxford : Clarendon Press, 1941. Price 8*s.* 6*d.* net.)

We have here, in surprisingly small compass, an admirable survey of the development of thought in the physical and biological sciences, up to the enthronement of the doctrines of energy, atomism and evolution. Beyond this stage it would scarcely be possible to go in a single volume addressed largely to non-specialists. Even within the limits so indicated, it would obviously have been far less difficult to write the same story in a thousand than in four hundred pages. Dr. Singer has achieved a masterpiece of graceful compression, and he has done a great service in producing a volume so handy, so serviceable and so salutary in these troubled times.

The treatment is scholarly without ostentation, and accurate without pedantry, though the rendering of Hooke's law as "ut torsio sic vis" (p. 304) is perhaps a little misleading, as a somewhat violent telescoping of Hooke's and Coulomb's contributions to the science of elasticity.

Dr. Singer has not merely succeeded in the difficult task of handling in one volume the complicated threads of many widely differing bodies of knowledge ; he has, in addition, woven these threads into a remarkably coherent pattern, and it may reasonably be hoped that many of his new readers will be brought to realize, or to appreciate more fully, how absorbingly interesting the history of science may be.

It would be difficult to praise this work too highly. It may perhaps be most appropriately commended for the qualities ascribed by its author to Archimedes—humanity, width of interest, simplicity of exposition and unity of purpose. H. R. R.

Diffusion in and through Solids. By Richard M. Barrer. [Pp. xiv+464.] (Cambridge : at the University Press, 1941. Price 30*s.*)

Dr. Barrer has done a service to physical chemistry in collecting in book form the widespread data on diffusion processes in solids. But the usefulness of this book is much more than that of a reference book for data. The author has discussed the diffusion processes in various types of system ranging, for example, from the diffusion of gases through metallic and non-metallic crystals and the interdiffusion of metals to the diffusion of gases through non-elastic polymers. These processes are all discussed in terms of the same theoretical background, and this leads to the co-ordination of the manifold phenomena of diffusion into a coherent whole. As the author states in his Introduction, diffusion processes are a branch of chemical kinetics, and in recent years great advances have been made in our understanding of diffusion processes in terms of molecular forces and configuration. On the other hand, diffusion processes are closely related to the rate and equilibrium conditions of surface adsorption and desorption, and the author has succeeded admirably in fusing together these points of view.

Of especial interest are those sections in which the author himself has made notable contributions both to the experimental and theoretical branches of the subject. Barrer's work on the sorption and diffusion of gases in zeolite structures and the diffusion of gases through polymeric substances is well known, but it of value to have this work set in the background of diffusion processes in general.

It is not only from the theoretical point of view that this book has value. nor is it of interest only to the pure physical chemist. The methods of experimentation are critically reviewed. Many of the problems discussed such as, for example, gas flow through metals, the permeability of refractories and the permeability of natural and synthetic polymers, are all topics which are of importance to the industries.

This book, I feel, will serve a dual purpose, on the one hand that of stimulating interest and further research into this fascinating topic, and on the other hand that of giving those who meet problems of this type in their work the necessary background for an understanding of the mechanism of the processes involved. M. G. E.

The Physical Examination of Metals. Vol. II. *Electrical Methods.* By BRUCE CHALMERS and A. G. QUARRELL. [Pp. viii+280.] (London: Edward Arnold & Co., 1941. Price 20*s*. net.)

IT is pleasant to find that war-time conditions have not prevented the appearance of Volume II. oft his useful work. Volume I., published a year or so ago, dealt with optical methods of examining metals. The present volume describes magnetic and electrical methods under the following headings :—magnetism, electrical measurements, X-ray diffraction, electron diffraction, the electron microscope and radiography. There is also an appendix of useful data. The book is written primarily for those concerned with the practical examination of metals, and the emphasis throughout is on the applied side of the subject. The authors have collected and classified under the above headings a wide range of practical applications which, together with the optical methods described in Volume I., are likely to be of the greatest interest to all workers on metals. The book will also be of service to teachers of physics in universities and technical colleges ; they will find a wealth of interesting material to supplement the usual discussions of basic principles.

The absence, for the most part, of references to sources of general information appears to the reviewer rather unfortunate, particularly since the discussion of fundamental principles and theories is necessarily brief. Physicists will doubtless know where to obtain such additional information, but other readers, metallurgists, engineers, chemists and others who are less well acquainted with physical literature, would probably welcome some guidance. The deficiency is partially satisfied by the numerous references to particular applications, but nevertheless the reviewer would suggest that when a second edition is required, a short bibliography to general literature should be added to each chapter.

The book is clearly printed on good quality paper, with numerous line diagrams and eight plates. Having regard to present conditions, the price of the book is reasonable and for its size it is no more expensive than Volume I. (Vol. I., 181 pages, 14*s.*; Vol. II., 280 pages, 20*s.*) G. W. B.

The Second Yearbook of Research and Statistical Methodology Books and Reviews. By O. K. Buros. [Pp. x+383.] (The Gryphon Press, New Jersey, 1941. Price $4.50.)

This is one of a series of publications which dealt at first with the literature of mental tests only, but expanded into a wider field. It gives a list (in alphabetical order of authors) of three hundred and fifty-nine books and pamphlets, accompanied by extracts from reviews; this occupies three hundred and fifty-six pages, and there are some auxiliary lists and a brief subject-index. The main list contains items on mathematical statistics, market research, history of science, agricultural research and so on, all in English, but the hope is expressed of wider extension to other languages and to periodical publications.

Thus the book is mainly a collection of reviews, reduced by excision. The purpose, clearly stated, is to make better known the wide variation of standard in books on the same subject (particularly statistical methods) with a view to discouraging the publication of rubbish, and at the same time to encourage more honest and better informed reviewing. These are excellent aims, but there is practically no evidence of them in the present volume. The conjunction of reviews still leaves the information as to contents in its usual inadequate and ambiguous condition, and produces no effect at all of the nature of selective criticism. There appears, indeed, to be no attempt to choose extracts to suit this purpose; much futile padding is included and publications which normally reach only a limited circle, and so are liable to be overlooked, get a bare mention, while others which have been widely discussed and are often referred to receive the bulk of attention. Thus, Bernal's 'Social Function of Science' gets nearly twelve pages of extracts, and under Burt's 'Factors of the Mind' there is reprinted nearly three pages of a review by Godfrey Thomson, valuable in itself but having no bearing on the position of the book in a reference list.

Of course there is always interest and often value to be found in reading reviews and this is a finely bound and printed collection, but then why spoil the flavour by giving excerpts only? Those who wish to encourage the compiler to continue can discuss details, but there is no sign here of any advance beyond, or even useful addition to, the lists and abstracts to which we are accustomed. G. S.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

IX. *The Thermal and Electrical Properties of Bitumastic Compounds containing Quartz Sand.*

By Prof. WILLIS JACKSON, Electro-Technics Department,
Manchester University *.

[Received December 3, 1941.]

Introduction.

DURING the immediate pre-war years considerable attention was given in Germany, in particular by Dr. Meissner † of A.E.G., to the problem of improving the thermal conductivity of electrical insulating materials. The insulating materials possessing the highest thermal conductivity are crystals, and of this class of substance the most readily available, cheaply and in a form suitable for direct application in the above connexion, is quartz. Crystalline quartz has a thermal conductivity 50 to 100 times higher than that of the average amorphous type insulating material. It occurs naturally, in a reasonably pure state, as sand, in which form it can be incorporated without difficulty in many amorphous materials. Meissner showed that the thermal conductivity of bitumastic type compounds in particular is increased about six times by the incorporation of 70 to 80 per cent. of quartz sand, without appreciable detriment to the electrical properties. The use of these latter mixtures to eliminate enclosed air spaces in certain pieces of electrical equipment led to a considerable saving of copper, and in consequence marked reductions in overall dimensions, as compared with older designs of analogous rating and performance. In view of the possibility that this development may have been overlooked by many in this country. and that there may now exist cases in which such mixtures might be used to advantage, it is thought worth while to record a series of measurements carried out four years ago by way of confirmation of Meissner's work.

Physical Properties of the Compounds used in the Investigation.

Three bitumastic compounds of appreciably different dropping point and pouring temperature were chosen; their respective properties are given in Table I.

---

* Communicated by the Author.

† A. Meissner, *E. T. Z.* lv. pp. 1193-1195, Dec. 6; pp. 1218-1222, Dec. 13. 1934. *E. u. M.* liii. pp. 289-293, June 23, 1935.

Grain Size of the Quartz Sand Employed.

Three samples of sand, designated by the suppliers as Fontainebleau, Belgian and German sand, were employed, and the analysis of their constituent grain sizes is given in Table II.

An isolated measurement was made on a mixture containing ground silica of grain size about 60 mesh.

TABLE I.

Compound.	Dropping point.	Pouring temperature.
	°C.	°C.
A	67·5	120
B	85·0	137
C	106·0	150

TABLE II.

Type of sand.	Per cent. residue of weight left in succession on various sieves.					
	40 mesh (0·63 mm.)	60 mesh (0·425 mm.)	80 mesh (0·32 mm.)	100 mesh (0·25 mm.)	120 mesh (0·21 mm.)	Residue.
Fontainebleau	0·0	2·2	49·5	38·8	3·9	5·6
Belgian	4·4	21·8	47·3	18·4	2·7	5·5
German	0·0	6·2	44·7	42·4	4·7	2·0

Results of Thermal Resistivity Measurements.

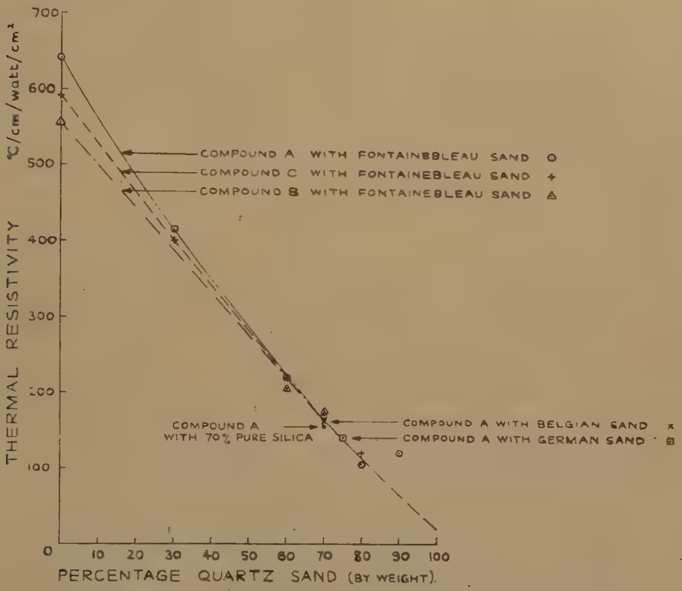
Measurements of thermal resistivity were made on mixtures of the compounds and sands mentioned in proportions from 0 to 90 per cent. by weight of sand. The test-specimens consisted of disks of 1½ in. diameter produced inside bakelite disks of ⅛ in. depth. The procedure in producing these disks was as follows :—

The compound was heated to a semi-fluid condition and the appropriate proportion of sand introduced in stages, care being taken to avoid the inclusion of air-bubbles during the mixing process. A bakelite ring, laid on the amalgamated surface of a hot brass plate, was filled with a slight excess of the fluid mixture and the excess removed by applying pressure from above by means of a second brass plate. The plates and the specimen were then cooled rapidly to avoid appreciable settling of the sand.

The tendency of the sand to settle, though readily overcome in the case of such small specimens, is an important consideration when large bulks of these mixtures are to be handled. The effect was not found to be a serious one, however, for quartz percentages of 60 or more, the order of sand addition to which most interest attaches, provided that the pouring temperature was kept as low as possible and the mixture kept well stirred until pouring was commenced.

The result of thermal resistivity measurements on a variety of mixtures is depicted in fig. 1. The conclusions to be drawn from the measurements are as follows :—

Fig. 1.



Dependence of thermal resistivity on percentage of quartz sand addition.

(1) The three compounds A, B and C have different thermal resistivities, 640, 555 and 590 units (°C./cm./watt/cm.^{2}) respectively, but for quartz sand additions beyond about 50 per cent. this difference exerts no important effect on the thermal resistivity of the mixture.

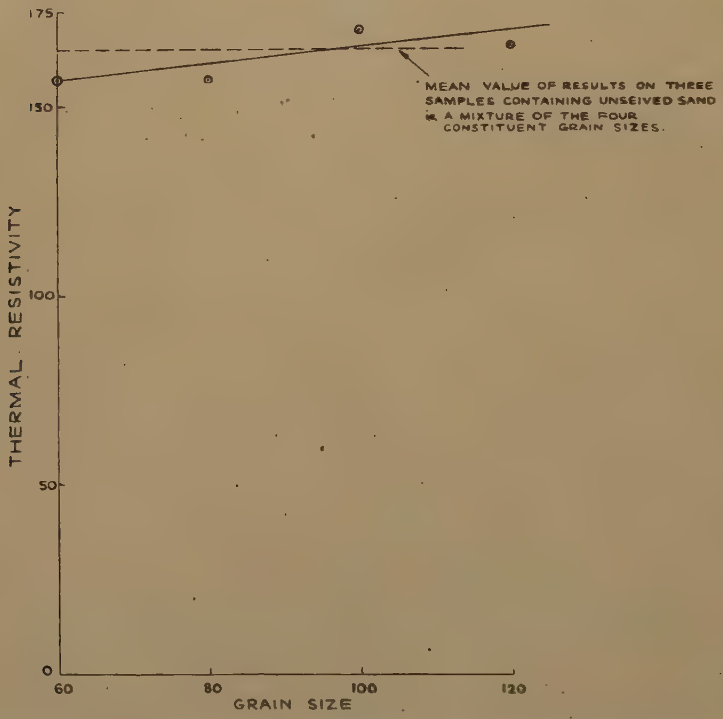
(2) The improvement produced by the addition of sand is sensibly the same for each of the three grades of sand mentioned, and also for pure silica.

(3) The addition of 80 per cent. sand by weight reduces the thermal resistivity to less than one-sixth of the value for the compound alone. The high figure given by a sample containing 90 per cent. sand suggests

that the amount of compound available was insufficient to fill completely the spaces between the sand particles.

(4) Extrapolation of the experimental curves to an imaginary condition of 100 per cent. by weight of quartz (a solid quartz disk), suggests a figure for the thermal resistivity of about 20 units. which compares favourably with the value 10 to 15 units usually associated with pure crystalline quartz.

Fig. 2.



Dependence of thermal resistivity on grain size for compound A containing 70 per cent. quartz sand by weight.

(The figure, 80 say, corresponds to the portion of the sand which passes through the 60 but remains on the 80 mesh sieve.)

Direct Investigation of the Effect of Grain Size on the Thermal Resistivity.

Fig. 2 shows the dependence of the thermal resistivity of a mixture of compound A and 70 per cent. by weight of sand particles on the grain size of these particles. The values of grain size given along the abscissa are averages in the sense that the figure of 80, for example, corresponds to the portion of the sand (Fontainebleau) which passed through a 60 but remained on an 80 mesh sieve. It is seen that the thermal resistivity varies by less than 10 per cent. over the range of grain size from 60 to 120, a result which is in keeping with that, referred to above, relating to the three different grades of sand.

Thermal Resistivity Measurements on Quartz Sand alone.

These measurements were made in anticipation of possible applications of the sand alone in cases where the operating temperature may be too high to permit the use of compound-quartz mixtures. The dependence of the thermal resistivity on the grain size, for samples subjected to the same degree of compression, was found to be as follows :—

	Thermal resistivity with		
	Zero pressure.	4 lb./in.^{2} pressure approx.	10 lb./in.^{2} pressure approx.
Unsieved sand (Fontainebleau)	368	358	341
60 mesh residue	374	335	314
80 mesh residue	353	341	328
100 mesh residue..................	323	309	297

These results are not so consistent as those shown in fig. 2, but again the variation of grain size over the range studied does not exercise a very important effect. The important observation is that the thermal resistivity is much higher than for comparable compound-quartz mixtures. Thus the compounds containing 80 per cent. by weight of quartz gave a thermal resistivity figure of only 110. This difference is not surprising, however, because of the much greater resistivity of the tiny air spaces present within the compressed sand as compared with that of compound filled spaces.

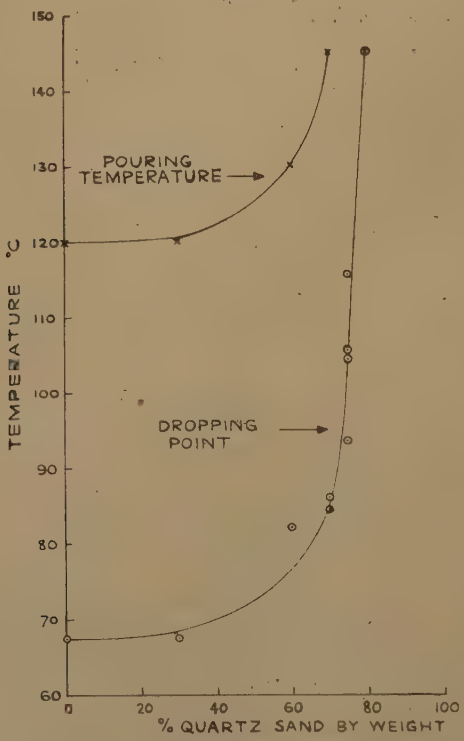
An attempt to reduce the thermal resistivity of these spaces by filling them with fine powders was not successful. This is shown in the following results :—

	Thermal resistivity. 4 lb./in.^{2} pressure approx.
Fontainebleau sand mixed with 10 per cent. by weight of micronized sand	330
Fontainebleau sand mixed with 10 per cent. by weight of fine asbestos powder......................	408
Fontainebleau sand mixed with 10 per cent. by weight of fine feldspar powder	442

Dropping Point and Pouring Temperature Measurements on Compound-Quartz Mixtures.

Fig. 3 shows the manner in which the dropping point and pouring temperature varied with the percentage of quartz by weight included in compound A. It is seen that the dropping point begins to rise very rapidly when the quartz sand percentage reaches about 70. This is emphasized by the fact that the figures obtained for four samples taken from a 75 per cent. mixing were 93·5, 104·5, 105·5, and 115·5° C. This variation may well be accounted for in terms of a small variation in the amount of quartz sand present in the respective samples ; it is unlikely that the samples compared to better than ±1 per cent. in this respect.

Fig. 3.



Dependence of dropping point and pouring temperature on percentage of quartz sand in compound A.

The pouring temperature appears likely to follow a similar course ; measurements were not carried out, however, beyond a 70 per cent. sand addition, since it would not be desirable in practice to heat this particular compound to temperatures in excess of 150° C.

Temperature Rise of Coils Surrounded by Various Media.

In order to present a more direct view-point on the thermal effect of the quartz sand addition, measurements were made of the temperature rise of four identical coils when supported centrally in metal containers with their axes vertical and surrounded respectively by air, A compound alone, A compound containing 70 per cent. by weight of Fontainebleau sand, and transformer oil. Current was passed through the coils in turn and the power dissipation made the same in each case. The temperature rise was derived from the increase in resistance of the coil.

Dimensions of coils.

Outside diameter	2·9 in.
Inside diameter	2·0 in.
Length.....................	2·1 in.

Dimensions of metal container.

5 in. high × 4·5 in. diameter.

Results.

	Power dissipated. Watts.	Air temperature. °C.	Temperature rise of coil. °C.
(1) Coil in empty container......	6·3	22·0	49·3
(2) Coil in compound A	6·3	22·0	36·5
(3) Coil in loaded compound A ...	6·3	22·0	23·6
(4) Coil in transformer oil	6·3	21·5	18·7

The loaded compound gave a marked improvement over the unloaded material, and in particular over the air enclosure, but was not so efficient as transformer oil in reducing the final temperature rise. There are many cases, however, where the latter would be unsuitable for use and where the loaded compound might be applicable.

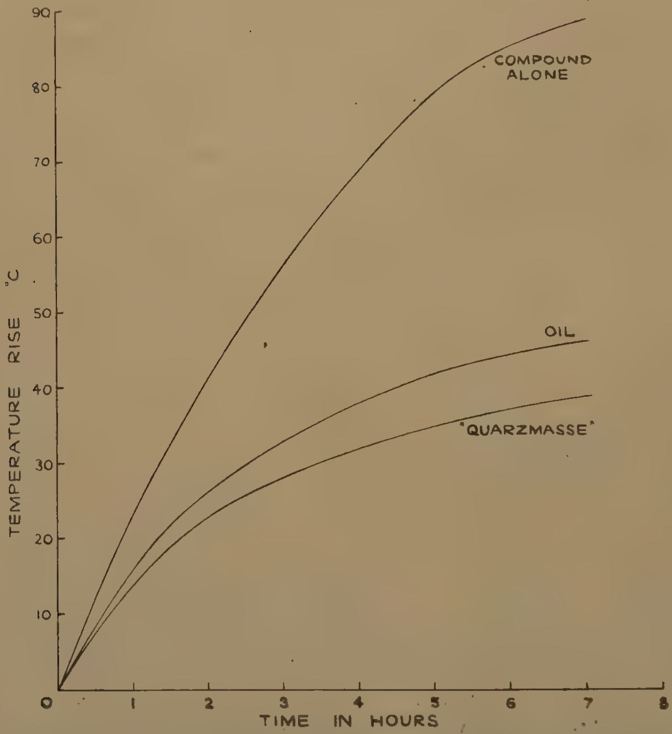
The curves of fig. 4, supplied by Meissner, give the results of temperature-rise measurements on a particular 0·5 kV.A. transformer, overall dimensions 22 × 17 × 14 cm., when the coils were surrounded by compound, oil and his " quartzmasse " respectively.

Electric Strength Measurements at 50 Cycles per Sec.

It is of interest to ascertain the extent to which the incorporation of quartz sand affects the electrical properties. The excellent dielectric properties of quartz suggest that the effect need not be seriously detrimental, and this is borne out by the experimental results given below.

Experiments were conducted only on compound A; first on the compound alone, and then on a mixture containing 70 per cent. by weight of Fontainebleau sand. The test-specimens consisted of disks of 6 in. diameter and $\frac{1}{8}$ in. thickness, prepared in the manner already described, and were tested between plane electrodes ($\frac{3}{4}$ in. and $1\frac{1}{2}$ in. diameter respectively) immediately after immersion in oil at 15° C. This procedure is not the standard one for the high voltage testing of insulating compounds, but was adopted as the most convenient means of obtaining the desired comparative information.

Fig. 4.



Temperature rise of a 0·5 kV.A. transformer filled with compound, oil and "quartzmasse" respectively.

The serious effect which air-pockets would have on the electric strength of the specimens was self-evident. Two sets of test-specimens were therefore prepared; one in which no special precautions were taken to prevent the existence of small air-pockets, and the other in which the prepared disks were re-heated under vacuum to remove occluded air.

The sand employed was maintained at 150° C. for several days before mixing.

The experimental results obtained were as follows :—

(1) *Compound alone* (test-specimens not heated under vacuum).

Electric strength figures	435, 540, 580, 440 and 560 volts/mil.
Mean value	510 volts/mil.

(2) *Mixture containing* 70 *per cent. quartz sand by weight.*

(*a*) Test-specimens not heated in vacuum	435, 330, 420 and 360 volts/mil.
Mean value	390 volts/mil.
(*b*) Test - specimens heated in vacuum	420, 400, 420, 410 volts/mil.
Mean value	410 volts/mil.

The improvement resulting from evacuation is not marked. The important observation is that although the electric strength is reduced by the addition of the quartz sand, it is still adequate for the purposes to which compound-quartz mixtures might be applied.

Power Factor Measurements.

These measurements were carried out at 50 cycles per sec., and room-temperature on specimens of the same dimensions as those employed in the breakdown tests. The results obtained were as follows :—

Compound A alone	1st specimen—0·016. 2nd specimen—0·017.
Compound A containing 70 per cent. Fontainebleau sand as received	1st specimen—0·038. 2nd specimen—0·046.
Compound A containing 70 per cent. Fontainebleau sand after washing in dilute HCl and distilled water, and subsequent drying at 150° C. for 7 days.	1st specimen—0·025. 2nd specimen—0·037. 3rd specimen—0·038.

It is seen that the quartz addition is responsible for an increase in power factor of roughly 100 per cent., even after the sand had been specially cleaned and dried. This increase is probably due to the presence of absorbed moisture on the surface of the sand, which even prolonged heating at 150° C. appears incapable of removing.

The author is indebted to Dr. A. P. M. Fleming, of Messrs. Metropolitan Vickers Electrical Co., Ltd., for permission to make this information generally available.

X. *Some Comments on the Article by C. G. Pendse.* " LXXXVII. *A Note on the Definition and Determination of Mass in Newtonian Mechanics.*"—Phil. Mag., Nov. (Suppl.) 1937.

By Glen H. Draper *.

[Received October 29, 1940.]

Frequently mathematicians have followed a definite procedure in their work, forgetting that the procedure is itself a major premise which should not be overlooked in the deductions, which have on occasion been quite absurd. By the use of an esoteric and abstract language the rest of mankind have been inveigled to accept the invalid conclusions. A case in point is the series of articles which have been published bearing on the subject of the one here discussed.

A deduction taken from page 1020 is quoted as follows :—" (3) The solar system may also be regarded as an ' isolated ' system ; but, strictly speaking, a Newtonian observer for the system would be unable to determine *uniquely* and *exactly* the ratios of the masses of the members of the system (which contains more than seven bodies) by observing their motions. He has to assume, as a first approximation, that the sun and one of the planets form an ' isolated ' system, and has to use the law of gravitation formulated from the orbit of the planet. He has to create mentally artificial ' isolated ' systems as a first approximation.

The procedure is permissible only because it is fortunately possible to ignore the interactions between the planets in comparison with the gravitation of the sun."

This assertion seems to say, one can solve a set of (n) simultaneous equations involving $(n+k)$ unknowns by approximations, providing one unknown is large in comparison with all the others (nonsense).

The fallacy in the conclusions of Dr. Pendse is readily perceived by changing the procedure of his analysis.

He postulates the $(3n)$ equations :—

$$\ddot{x}_i=\ddot{x}_{i0}+\dddot{x}_{i0}\,.\,t+\tfrac{1}{2}\ddddot{x}_{i0}\,.\,t^2+\,\ldots\,;\;\; x_i,\, y_i,\, z_i\,.\,i=1,\,2,\,\ldots,\,n$$

as being precisely known, see page 1017,—" consequently the observer is supposed to know the motion of every particle in the system . . .

---

* Communicated by the Author.

(at as many instants as desired)." There is nothing implicitly or explicitly forbidding knowledge by observations of the approximate values of x_i, y_i, z_i, in fact they are implicitly assumed to be known as they are the quantities observed and from which the accelerations are derived. He accepts the Three Laws of Motion, the corollary, and the law of resolving vectors into component parts. Hence, we may write $(3n)$ equations immediately:—

$$\ddot{x}_i = \sum_{j=1}^{n} \frac{f_{ij}}{m_i}\left(\frac{x_j - x_i}{\rho_{ij}}\right);$$

where $$f_{ij} = -f_{ji} = m_i m_j f(\rho, \dot{\rho}, \ddot{\rho}, \ldots).$$

Only $(3n-3)$ of which are independent, and because of the assumptions concerning x_i, y_i, z_i the only unknowns are (f_{ij}/m_i).

And if (f_{ij}) is assumed, then there remain but $(n-1)$ independent unknowns, viz. (m_i/m_1).

Since only $(n-1)$ independent equations are required to determine uniquely and exactly these unknowns, there remain $(2n-4)$ equations which are used to determine the correctness of the assumptions:—(1) "isolated" system, (2) homogeneous spherical bodies, which are necessary to be assumed to be particles, (3) the law (f_{ij}), and possibly (4) the laws of motion.

Therefore, we see that the procedure of astronomers in approximating the ratios of the masses in the solar system is due, not to an intrinsic difficulty, but solely to the lack of knowledge of the quantities $\ddot{x}_i$.

Many of the other statements which should be criticised are dependent upon the foregoing and hence not mentioned here, but another independent "esotericism" should not be overlooked. On page 1016 the symbol (m_{ij}) is defined to be (m_i/m_j), the arithmetical ratio; then on page 1021 in appendix the rule $m_{13} = m_{12} \times m_{23}$ is called an independent axiom, and "is to be verified for all groups of three particles." A definition of a symbol cannot be verified, and because of the definition the rule is precise. *One can only verify that he uses the symbol consistently with the explicit definition.*

XI. *Observations on Mr. Glen H. Draper's Paper entitled some Comments on the Article by C. G. Pendse* " LXXXVII. *A Note on the Definition and Determination of Mass in Newtonian Mechanics.*"—Phil. Mag., Nov. (Suppl.) 1937.

By C. G. PENDSE *.

[Received November 11, 1941.]

(I.)

THE three papers of mine on the subject of Mass in Newtonian Mechanics, published in the ' Philosophical Magazine,' Series 7, are :—

(i.) " A Note on the Definition and Determination of Mass in Newtonian Mechanics," vol. xxiv. pp. 1012–1022 (1937).

(ii.) " A Further Note on the Definition and Determination of Mass in Newtonian Mechanics," vol. xxvii. pp. 51–61 (1939).

(iii.) " On Mass and Force in Newtonian Mechanics—Addendum to ' Mass I.' and ' Mass II.'," vol. xxix. pp. 477–484 (1940).

They are referred to as " Mass I.", " Mass II.", and " Mass III." respectively.

Mr. Glen H. Draper, whose paper has been evoked by " Mass I.", has made no reference to " Mass II." and " Mass III." From " Mass I." he has selected example (3), concerning the solar system, on p. 1017 for analysis, and has briefly considered the statement that the rule

$$m_{13}=m_{12}\times m_{23}$$

is an independent axiom on p. 1021.

(II.)

It will be useful to recapitulate the relations between " Mass I.", " Mass II.", and " Mass III."

In addition to stating further developments, " Mass II." contained clarification and modification of some results in " Mass I."

In what follows, the undermentioned abbreviations are used :—

Case I. = Case (I.) on pp. 1017–1019 of " Mass I."
Case II. = Case (II.) on p. 1019 of " Mass I."

---

* Communicated by the Author.

Heuristic aspect = The heuristic aspect discussed in §3, pp. 55–57, of "Mass II."
Computational aspect = The computational aspect discussed in §4, pp. 58–60, of "Mass II."

It is seen from "Mass II." that the heuristic aspect and the computational aspect correspond to Case I. and Case II. respectively.

In both the heuristic aspect and Case I. the ratios of the masses of particles (and the forces between them) are indeterminate when their number exceeds four. But, whereas in Case II. the ratios of masses as well as the interactions remain indeterminate when the number exceeds seven, according to the computational aspect, the forces remain indeterminate when the number exceeds seven, whereas the ratios of masses *might* be determinable from the observed motions when the number exceeds seven: this is the chief modification.

The examples, particularly (3) and (4) on pp. 1020–1021 of "Mass I.", have to be considered after taking into account the results in "Mass II.", in which they were not discussed again as the modifications appeared to be obvious. For a restatement of these examples, in the light of the results in "Mass II.", it would be necessary to have two statements for each of the two examples, one statement corresponding to one aspect. The statements would be as follows:—

Example (3), *Heuristic aspect*:—As the solar system contains more than four bodies, it is not possible for an observer to determine *uniquely* and *exactly* the ratios of their masses or the interactions between them from a complete knowledge of their motions. He has to create mentally artificial systems of two bodies as a first approximation, the sun being one member and a planet another in each system. If he satisfies approximately the tests for a Newtonian observer for each such pair, and asserts that it is "isolated," he can deduce approximately the ratio of masses for each pair and determine the law of force. Thus, by a process of approximation and subsequent induction, he can find the ratios of masses and the forces.

The procedure is permissible only because it is fortunately possible to ignore the interactions between the planets in comparison with the gravitation of the sun. This is indeed analogous to the manner in which the law of gravitation was formulated.

Example (3), *Computational Aspect*:—The solar system may also be regarded as an "isolated" system. A Newtonian observer for the system would be able to determine the ratios of the masses of the members of the system, but not the forces between them, in a *unique and exact manner* by observing their motions. He has to create mentally artificial systems of two bodies as a first approximation, the sun being one member and a planet another in each system, and by induction arrive at the law of gravitation.

The procedure is permissible only because it is fortunately possible to ignore the interactions between the planets in comparison with the gravitation of the sun.

Example (4), *Heuristic Aspect* :—Finally, we consider the case in which the masses of the particles, numbering more than four, in a system do not vary widely, as in the case of a mixture of gases *. If we are presented with the knowledge of the motion of each particle it would be impracticable for an observer even to verify whether or no he satisfies the criteria for a Newtonian observer, and it would be quite impossible for him to determine the ratios of the masses of the particles or the forces between them from a complete knowledge of the motions.

Example (4), *Computational Aspect* :—Finally, we consider the case in which the masses of the particles, numbering more than seven, in an "isolated" system do not vary widely, as in the case of a mixture of gases. If we are presented with the knowledge of the motion of each particle, it would be quite impossible for us to determine the forces between the particles, though it might be possible to determine the ratios of their masses from a complete knowledge of the motions †.

(III.)

We now come to Mr. Draper's analysis of Example (3).

It has been mentioned already that "Mass II." contained clarification and modification of "Mass I." as well as having additional results. The statements concerning Example (3) given above state the whole situation. I think that the computational aspect of Example (3) mentioned above effectively answers the objection which has been raised. It may be added that the astronomer is mainly concerned with the computational aspect.

Regarding Mr. Draper's interpretation of the expression "first approximation," it is desirable to say that I have used it more in the manner of Pearson * than in the sense of a mere algebraic manipulation.

---

* When we speak, in the two aspects of Example (4), of the masses of particles not varying widely, we have in mind a system, in the case of which we believe for some reason or another that the masses do not differ widely from each other. Thus, we are applying the purely theoretical standpoint, from which the question of mass has been considered, to systems, working information regarding which has been obtained by the practised experimental and observational methods.

† *Cf.* the example on pp. 59–60 of "Mass II."

In view of what has been said above, it is felt that it is scarcely necessary to write more about Mr. Draper's analysis.

Now I come to Mr. Draper's comment on the remarks on the rule $m_{13}=m_{12}\times m_{23}$ on p. 1021 of Mass " I." This point required elaboration, which was given in §4 of " Mass II." I may add that the $m_{\lambda\mu}$ considered by me is not merely the ratio of prescribable algebraic mass numbers. It is the negative inverse ratio of the mutually induced accelerations. The rule has been fully explained in " Mass II."

The object of the investigation in " Mass I." and " Mass II." has been again emphasized in " Mass III." I would like to mention particularly paragraph 4 of the Introduction, and the last two paragraphs of §5 of " Mass III."

---

* Karl Pearson, 'The Grammar of Science.' Chapter III. (" The Scientific Law"), §10 (" Progress in the Formulating of Natural Law"). (Everyman's Library edition, p. 85.)

XII. *Collision Relations.*

By L. A. Pars *.

[Received October 3, 1941.]

The term "collision" includes problems of two types. The first is the impact of smooth spheres—the billiard-ball problem. The second is the interaction between two particles under a mutual attraction or repulsion; the mutual potential energy V is a function of r, the distance separating the particles, and V tends to zero as r tends to infinity. The motion of each particle at the beginning of the interaction, and again at the end, is sensibly uniform and rectilinear, r being infinite at the beginning and at the end. The most important case is that of the collision of an α-particle with an atomic nucleus †; in this problem the second particle is initially at rest. It is important to observe that many of the classical results—equations (1) to (4) below—are independent of the particular law of force; it is only necessary to assume that V is a function of r which tends to zero as r tends to infinity. These results are familiar ‡, but there is an extremely simple and elegant way of proving and exhibiting them which perhaps deserves to be more widely known.

We begin with the elementary problem of smooth spheres in collision. We denote the velocity before impact of the first sphere, of mass m_1, by $\mathbf{U}_1$, and the velocity before impact of the second sphere, of mass m_2, by $\mathbf{U}_2$. The corresponding velocities after the collision are denoted by $\mathbf{V}_1$ and $\mathbf{V}_2$. We consider a diagram in which the vector $\mathbf{U}_1$ is denoted by OU_1, the vector $\mathbf{U}_2$ by OU_2, and so on. The vectors $\mathbf{U}_1—\mathbf{V}_1$, $\mathbf{V}_2—\mathbf{U}_2$ have the same direction, the direction of the line of centres at the instant of impact. This direction is taken horizontal in the diagram (fig. 1). The point G divides U_1U_2 in the ratio $m_2 : m_1$, and the vector OG represents the velocity of the centre of gravity—a velocity which is unaltered by the collision. The point P_1 is the foot of the perpendicular from G on to the line U_1V_1, and the point P_2 is the foot of the perpendicular from G on to the line U_2V_2. The vectors OP_1, OP_2 represent the velocities of the spheres at the instant of maximum compression. We notice that, in the general case, the origin O is not coplanar with the other points shown in the figure.

---

* Communicated by the Author.

† I wish to thank Dr. N. Feather for his helpful suggestions as to the presentation of the theory, which have enabled me to make this note more useful to the experimental physicist than it would otherwise have been.

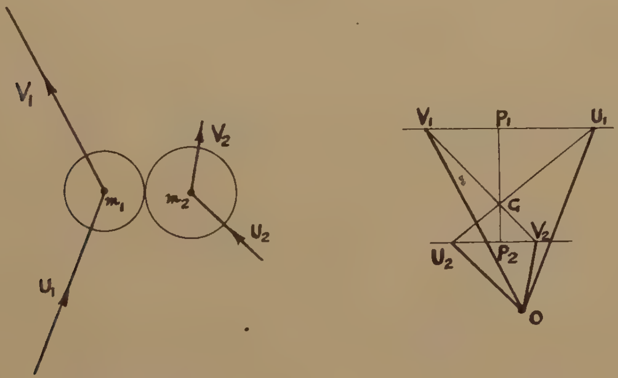
‡ See, for example, 'Radiations from Radioactive Substances,' Rutherford, Chadwick and Ellis, Cambridge, 1930, Chapter IX.

Now, referring to fig. 1, we have

$$\frac{V_1P_1}{P_1U_1}=\frac{P_2V_2}{U_2P_2}=\frac{V_1P_1+P_2V_2}{P_1U_1+U_2P_2}=e,$$

by Newton's law of impact, where e is the coefficient of restitution. This result includes the classical theorem "The increase of velocity of either sphere after maximum compression is e times the decrease of velocity before maximum compression." The theorem is an immediate consequence of the conservation of momentum in conjunction with Newton's law. We notice that energy is not conserved except in the case of a perfectly elastic impact, when $e=1$.

Fig. 1.



It is important to observe that, when U_1 and U_2 are given we can find V_1 and V_2 by an elementary construction. For, given U_1 and U_2, we first mark G, and then we find P_1 by dropping the perpendicular from G on to the line drawn through U_1 in the direction of the line of centres; the point V_1 is then determined by producing U_1P_1 to V_1, where

$$V_1P_1=e\,.\,P_1U_1.$$

A similar construction determines V_2.

The most important case is that in which the second sphere is initially at rest, and the impact is perfectly elastic, *i. e.*, $\mathbf{U}_2=0$ and $e=1$. In this case the energy is conserved, as we have already noticed. The figure now takes a simpler form. The point U_2 coincides with O, the figure is plane (the motion takes place in the plane determined by the vector **U**$_1$ and the line of centres), and there is symmetry about P_1P_2. Fig. 2 represents the case $m_1>m_2$, and fig. 3 represents the case $m_1<m_2$. The angles θ and ϕ are the inclinations of the final velocities of m_2 and of m_1 to the initial velocity of m_1.

The velocity of m_1 relative to G is unchanged in magnitude by the collision, and turned through an angle which we call 2ψ; moreover, the velocity of m_1 relative to m_2 is unchanged in magnitude and turned through the same angle 2ψ. The angles ψ and θ are complementary.

Fig. 2.

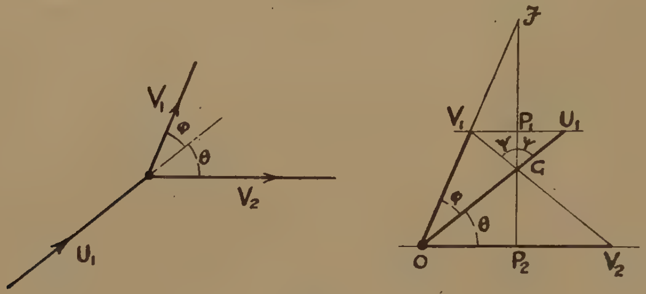
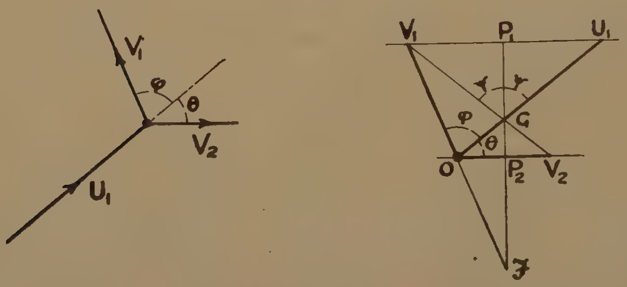


Fig. 3.



Now the same figures also exhibit the result of a collision of the second kind, where there are two particles under a mutual attraction or repulsion. The motion of m_1 relative to the centre of gravity is unchanged in magnitude by the collision, and turned through an angle which we call 2ψ. This angle 2ψ is also the angle through which the motion of m_1 relative to m_2 is turned by the interaction, and this angle can be calculated from the theory of orbits when the law of force is known. However, as we have already remarked, many of the important results are independent of the particular law of force.

If we now take P_1 and P_2 as fixed points, we can include the theory for different values of ψ in the same figure. (Of course if **U**$_1$ is thought of as of fixed magnitude, this involves the use of a different velocity-scale for different values of ψ, but this is no disadvantage, since only ratios of velocities are involved.) The point F is fixed as well as G, since F and G divide P_1P_2 harmonically, and it is easy to see how ϕ varies with ψ (or with θ). If $m_1 > m_2$ (fig. 2) ϕ increases steadily from zero to a maximum, and then decreases steadily to zero, as ψ increases from 0 to $\frac{1}{2}\pi$ (*i. e.*, as θ decreases from $\frac{1}{2}\pi$ to 0). The maximum value of ϕ occurs when the circle FGO touches OP_2V_2 at O ; in that case the angle GFO is equal to θ, GV_1 is perpendicular to V_1F, and

$$\sin\phi = \frac{V_1G}{GO} = \frac{m_2}{m_1}.$$

If $m_1 < m_2$ (fig. 3) ϕ increases steadily from 0 to π as ψ increases from 0 to $\frac{1}{2}\pi$ (*i. e.*, as θ decreases from $\frac{1}{2}\pi$ to 0).

We can now read off from the figure the various relations which are needed in the theory of the collisions of α-particles with atomic nuclei or electrons. The fundamental relation is that giving the ratio of the masses in terms of the angles θ and ϕ ; this relation is of great value in interpreting measurements of the branched tracks observed in the expansion chamber. We have immediately, on reference to fig. 2 or fig. 3,

$$\frac{m_2}{m_1} = \frac{V_1G}{GV_2} = \frac{V_1G}{GO} = \frac{\sin\phi}{\sin(2\psi-\phi)} = \frac{\sin\phi}{\sin(2\theta+\phi)}\text{[*]}. \quad . \quad . \quad . \quad (1)$$

This result is also exhibited by Rutherford in three other forms. Solving the equation (1) for $\tan\phi$ we have

$$\tan\phi = \frac{m_2 \sin 2\theta}{m_1 - m_2 \cos 2\theta}, \quad . \quad . \quad . \quad . \quad . \quad (1\,a)$$

and solving for $\tan\theta$ we have

$$\tan\theta = \frac{m_2\cos\phi \pm \sqrt{(m_2{}^2 - m_1{}^2 \sin^2\phi)}}{(m_1+m_2)\sin\phi}\text{[†]}; \quad . \quad . \quad (1\,b)$$

if $m_1 > m_2$ both signs are admissible before the radical, giving the two values of θ corresponding to any value of ϕ less than $\sin^{-1}(m_2/m_1)$; if $m_1 < m_2$ the plus sign must be taken. The results (1 *a*) and (1 *b*) can also be found. if preferred, directly from the elementary geometry of the figure. The last of the three forms is got by writing equation (1 *a*) in the form

$$\cot\phi = \frac{m_1 + m_2\cos 2\psi}{m_2 \sin 2\psi} = \cot 2\psi + \frac{m_1}{m_2}\operatorname{cosec} 2\psi\text{[‡]}. \quad . \quad . \quad (1\,c)$$

---

[*] Rutherford, *ibid.* p. 246.

[†] Rutherford, *ibid.* p. 241, equations (2) and (3).

[‡] Rutherford, *ibid.* foot of p. 242 ; the angle 2ψ is there denoted by ϕ'.

Next we require the formulæ for the final velocities of the two particles in terms of the initial velocity of m_1 and the observed angles θ and ϕ. We denote $|\,\mathbf{U}_1\,|$ by u_1, and so on, so that in the figures u_1 is the length OU_1. We have immediately

$$v_2 = OV_2 = 2OG\cos\theta = 2\frac{m_1}{m_1+m_2}u_1\cos\theta, \qquad (2)$$

and

$$v_1 = OG\cos\phi \pm \sqrt{\{(GV_1)^2-(OG)^2\sin^2\phi\}}$$
$$= \frac{m_1\cos\phi \pm \sqrt{(m_2{}^2-m_1{}^2\sin^2\phi)}}{m_1+m_2}u_1\text{[*]}. \qquad (3)$$

Equation (2) allows us to calculate immediately the fraction of the kinetic energy which is transferred from the projected particle to the particle initially at rest. If the energy transferred is E_2 and the initial energy is E_0 we have, from equation (2),

$$\frac{E_2}{E_0} = \frac{m_2v_2{}^2}{m_1u_1{}^2} = \frac{4m_1m_2}{(m_1+m^2)^2}\cos^2\theta, \qquad (4)$$

giving E_2/E_0 in terms of the angle θ and the ratio of the masses. It will be seen that the fraction of the energy transferred is a maximum when the first particle is projected directly towards the second, $\theta=0$, and the fraction transferred decreases steadily as θ increases from 0 to $\frac{1}{2}\pi$. The maximum value of the fraction transferred is

$$\frac{4m_1m_2}{(m_1+m_2)^2} = 1 - \left(\frac{m_1-m_2}{m_1+m_2}\right)^2,$$

which has a maximum value 1 when $m_1=m_2$; in this case *all* the energy is transferred in direct impact, as is well known. In the extreme case $m_1 \gg m_2$ (as in the collision of an α-particle with an electron) the maximum transferred is approximately $4m_2/m_1$.

It is most important to observe that all the preceding results are independent of the particular law of attraction or repulsion between the particles.

Finally, if we now introduce the assumption of a particular law of force between the particles, we can determine the angle ψ in terms of the initial conditions of the experiment. By a well-known theorem in the theory of orbits, the motion of m_1 relative to m_2 is the same as if m_2 were fixed, and the mass of m_1 decreased to ρ, where

$$\frac{1}{\rho} = \frac{1}{m_1} + \frac{1}{m_2}\text{[†]}.$$

---

[*] Rutherford, *ibid.* p. 241.

[†] Most authors are content to give the proof for the special case of the inverse-square law; see, for example, H. Lamb, 'Dynamics,' p. 235 (1930).

In particular, if the law of force is an inverse-square law, with an attraction or repulsion ZeE/r^2, it is well known that the relative orbit is a hyperbola, and that the angle between the asymptotes is 2θ, where

$$\tan\theta = pu_1^2/\mu\ ;$$

here p denotes the "impact parameter"—*i. e.*, the initial distance of m_2 from the initial line of motion of m_1—and $\mu = ZeE/\rho$. Thus

$$\tan\psi = \cot\theta = \frac{ZeE}{\rho}\,\frac{1}{pu_1^2} = \frac{1}{p}\,\frac{ZeE}{u_1^2}\left(\frac{1}{m_1}+\frac{1}{m_2}\right)^{*} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

---

* Rutherford, *ibid.* p. 242.

XIII. *Corresponding Problems in Periodic and Steady Flow.*

By George Green, D.Sc., Lecturer in Applied Physics
in the University of Glasgow *.

[Received October 17, 1941.]

In the periodic flow problems which we shall be considering we are concerned with solutions of the equations

$$\frac{d^2v}{dt^2}=V^2\nabla^2(v)\,; \quad \text{and} \quad \frac{dv}{dt}=\kappa\nabla^2(v)\,; \qquad (1)$$

the wave-equation and the general equation governing heat conduction respectively. When v varies with the time t according to the factor e^{ikt}, the equation to be solved in each case is

$$(\nabla^2+\lambda^2)v=0, \qquad (2)$$

where $\lambda^2=(k^2/V^2)$ in wave-motion problems, and $\lambda^2=-\frac{ik}{\kappa}$ in heat conduction problems.

When the problem to be considered relates to steady flow of fluid or of heat, the fundamental equation to be solved is that obtained from (1) and (2) by suppressing the term involving the time, namely

$$\nabla^2(v)=0. \qquad (3)$$

The present investigation is a preliminary study of the relations holding between solutions of (1) and solutions of (3) in a group of physical problems.

In general, any steady flow problem can be solved directly by determining the solution of equation (3) which satisfies prescribed boundary conditions. On the other hand, we shall find that it is possible to solve the same steady flow problem indirectly by first determining the solution of the corresponding periodic flow problem in which the flow takes place under the same boundary conditions. The process whereby we may obtain the required solution for the case of steady flow from the periodic solution obtained relating to the same boundary conditions depends upon a property of the Fourier integral

$$I=\frac{1}{\pi}\int_0^\infty dk\cos\{k(x\mp Vt)\}\,; \quad V=f(k)\,; \qquad (I.)$$

---

* Communicated by the Author.

which appears in Fourier's double integral theorem. Taken in relation to wave-motion and transmission of effects through a uniform medium, this "instantaneous integral" represents an effect at place x and at time t transmitted by wave-trains from a localized finite action or effect instantaneously present and concentrated at the origin at $t=0$[*]. The instantaneous integral provides an expression representing a finite initial effect or quantity as a summation of periodic effects[†]. The application of this integral to heat conduction and other problems has been made in previous papers. It is convenient accordingly to illustrate the process by which steady flow solutions are obtained by reference to fundamental cases in the subject of heat-flow.

Numerous applications of the process referred to to particular problems have been made throughout previous papers by the present writer. Hitherto, however, the process has not been made use of as a general method capable of providing solutions to new problems. The fundamental idea used in the present investigation is that every periodic solution (of (1)) can be used to provide a corresponding continuous solution (of (3))—by the process applied to various periodic source solutions only in previous investigations. As the notation employed below is that of previous papers to which frequent reference must be made, a list of the more important of these papers is given (p. 114).

In Part I. of the investigation given below, the main theorem required in the application of the process is demonstrated, and the method of obtaining solutions by its means is tested by applying it to problems whose solutions are known.

In Part II., applications are made to new problems.

PART I.

§1. *Flow in an Infinite Medium.*

Point Sources.—The fundamental periodic solution of equation (2) is

$$v=\frac{e^{ikt-i\lambda r}}{r}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

which, when λ is given a suitable value, is applicable to either of the equations appearing in (1). This solution in every case represents a point source. For example, the periodic heat source of strength qe^{ikt} situated at the origin is represented by

$$v_p=\frac{q}{4\pi K}\frac{e^{ikt-i\lambda r}}{r}: i\lambda=\sqrt{\frac{ik}{\kappa}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

---

[*] Proc. Roy. Soc. Edin. xxix. part v. pp. 445–470; xxx. part iii. pp. 242–253.

[†] Numerous examples of the application of the "instantaneous integral" occur in writings by the late Prof. Horace Lamb, F.R.S., on Water Waves: *e. g.*, Phil. Mag. xxxi. p. 386 (April 1916); and p. 539 (June 1916).

or by
$$v_p=\frac{qe^{ikt}}{4\pi K}\sqrt{\frac{2}{\pi}}\,.\,i\lambda\frac{K_{\frac{1}{2}}(i\lambda r)}{(i\lambda r)^{\frac{1}{2}}} \qquad (5')$$
in terms of the Bessel function ((5), p. 923).

The instantaneous integral, formed from this in accordance with (I.) above, is defined by
$$v_i=\frac{1}{\pi}\int_0^\infty dk\, v_p=\frac{q\kappa e^{-\frac{r^2}{4\kappa t}}}{8K(\pi\kappa t)^{3/2}}, \qquad (6)$$
which, as is well known, represents the effect of a finite heat quantity, q, instantaneously delivered to the medium at instant $t=0$, at the origin ((1), pp. 787, 794).

The continuous point source solution, representing the effect of a continuous heat supply at the origin of amount q per unit time, is obtained from the instantaneous point source solution v_i by the process indicated by
$$v_c=\underset{t=\infty}{\mathrm{Lt}}\int_0^t dt'\, v_i(t-t'). \qquad (7)$$
This process yields, for the value of v_i given in (6),
$$v_c=\frac{q}{4\pi K r}. \qquad (8)$$
The continuous flow solution (8) has therefore been derived from the periodic flow solution (5) by a process which involves only a simple application of Fourier's integral theorem, and which therefore may be employed to obtain from any periodic solution (of equation (1)) the corresponding continuous flow solution (of equation (3)), provided the limiting value represented by (7) can be found. The same process may be employed in the case of the point source solution of the wave-equation ((5), p. 924); the process is the same as that applied to the periodic point source in heat conduction, but the "instantaneous integrals" differ in the two cases.

It is obvious that a process which is valid when applied to point sources must also be applicable to distributions of point sources and also to multiple sources, with which we are concerned in applications made below.

Line Sources.—Corresponding to the point source solutions represented by (5) and (8) above, we have the line source solutions
$$v_p=\frac{qe^{kt}}{2\pi K}K_0(i\lambda\rho) \qquad (9)$$
and
$$v_c=-\frac{q}{2\pi K}\log\rho. \qquad (10)$$

The first represents a periodic line source of strength qe^{ikt} per unit length, and the second a continuous line source of strength q per unit length; and they are obtainable by integrating (5) and (8) respectively along an infinite line.

From (9) we obtain the instantaneous line source solution given by the Fourier instantaneous integral

$$v_i = \frac{q}{2\pi^2 \mathrm{K}} \int_0^\infty dk e^{ikt} \mathrm{K}_0(i\lambda\rho) \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

$$= \frac{q}{4\pi \mathrm{K} t} e^{-\frac{\rho^2}{4\kappa t}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11')$$

For this evaluation see ((2), pp. 705–6).

In this case also the application of the process indicated by (7) is consistent with the result (10). The results stated here involve the inherent difficulty that the integration along an infinite line introduces an infinite term, which is omitted from (10) as of no physical significance *. It is, in fact, the gradient of v_c that is of importance: to maintain that gradient at each point near the line source an infinity in the temperature values is required. A similar case of infinite temperature occurs in (17′) below.

Plane Sources.—There are two plane source solutions of special interest as illustrations of the method under consideration.

(*a*) The periodic solution

$$v_p = \theta_0 e^{ikt - i\lambda x} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

represents temperature at all points when the plane $x=0$ is maintained at temperature $\theta_0 e^{ikt}$. In accordance with the interpretation of the Fourier "instantaneous integral" given above, the effect due to the instantaneous creation of temperature θ_0, at the origin at instant $t=0$, is represented by

$$v_i = \frac{\theta_0}{\pi} \int_0^\infty dk \, e^{ikt - i\lambda x} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

$$= \frac{\theta_0 x e^{-\frac{x^2}{4\kappa t}}}{2\sqrt{\pi \kappa t^3}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13')$$

The continuous solution follows from this ((1), pp. 785–7), in accordance with (7), the required integration being a repetition of that performed in equation (8) above,—namely

$$v = \theta_0 . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

---

* Infinities occur also in all expressions representing concentrated initial effects, *e. g.*, instantaneous integrals.

Compare this with (5) and (8): t and k appear in both cases in the same way.

(*b*) The periodic solution

$$v_p=\frac{q}{2\mathrm{K}i\lambda}e^{ikt-i\lambda x} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

represents the effect at any point in an infinite medium due to a periodic heat supply qe^{ikt} per unit area of the plane $x=0$. In this case the Fourier "instantaneous integral" represented by (6) above yields the well-known instantaneous heat source solution

$$v_i=\frac{q\kappa e^{-\frac{x^2}{4\kappa t}}}{2\mathrm{K}\sqrt{\pi\kappa t}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

indicating the effect due to the instantaneous delivery of heat of amount q units per unit area of the plane, $x=0$, at instant $t=0$ ([1] pp. 786–7).

The effect due to a continuous heat supply at $x=0$ per unit area per unit time, as determined by (7) above, is given by

$$v_c=\frac{qx}{2\mathrm{K}\sqrt{\pi}}\cdot \underset{t=\infty}{\mathrm{Lt}}\int_z^\infty dp\, e^{-p^2}/p^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

where $z^2=x^2/4\kappa t$. An integration by parts yields the result

$$v_c=-\frac{q}{2\mathrm{K}}x+\mathrm{I}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17')$$

where I represents a divergent integral. Now it can be shown that the gradient of v_c does not depend on I, exactly as in equation (10) above. With a uniform gradient maintained throughout an infinite medium infinite temperature values must appear. On this understanding we take

$$v_c=-\frac{q}{2\mathrm{K}}x \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17'')$$

as the continuous flow solution corresponding to (15). Equation (17′) expresses the fact that a continuous heat supply at $x=0$ requires the establishment of a uniform gradient throughout the medium, when conditions of flow have become steady.

The two cases (*a*) and (*b*) refer respectively to temperature sources and heat sources in relation to plane flow of heat. It is to be noted, however, that each temperature source is simply $\left(-\frac{2\mathrm{K}}{q}\cdot\theta_0\frac{d}{dx}\right)$ of the corresponding heat source. This relation holds for temperature and heat sources in heat flow generally, the temperature source being in fact the doublet source with reference to the heat source ([6], pp. 936–7).

Theorem.—Consider now the "instantaneous integral" in any of the forms in which it appears above in plane, or cylindrical or radial flow, {(6). (11), (13), (16)}. From (16) we see that in plane flow the instantaneous integral, as a function of t, is of the form $\{e^{-c/t}/t^{\frac{1}{2}}\}$, and in the general case it is a function of the expression representing the instantaneous point source. Let the integration represented by (7) above be applied to the function

$$v_i = \frac{d}{dt}\{e^{-c/t}/t^n\} = \frac{d}{dt}f(t), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

assumed to represent an "instantaneous integral." The result obtained is $\underset{t=\infty}{\mathrm{Lt}}\,[f(t-t')]_0^t$, the value of which is zero. Similarly a zero value is obtained when the above and also other instantaneous integrals mentioned in ((5), pp. 931–4) are tested in the same way. The same result holds also for successive derivatives of these functions with respect to t.

The following theorem therefore applies to all periodic solutions yielding instantaneous integrals of the above types—namely, that for these solutions no corresponding steady flow solution exists. Taking, for example, periodic solutions derived from (12) and (15) by differentiation with respect to time, the theorem asserts that, corresponding to periodic solutions represented by

$$v_p = \lambda^n e^{ikt - i\lambda x}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (B)$$

no steady flow solution exists if $n=1, 2, 3$, etc.

As an illustrative test of the validity of this theorem we may examine its application to the group of periodic solutions represented by

$$v_p = \frac{e^{ikt}}{(ar)^{\frac{1}{2}}}\mathrm{K}_{n+\frac{1}{2}}(i\lambda r)\mathrm{I}_{n+\frac{1}{2}}(i\lambda a)\mathrm{P}_n(\cos\theta), \quad . \quad . \quad . \quad . \quad (19)$$

$$= \frac{e^{ikt}}{(ar)^{\frac{1}{2}}}p_{n+\frac{1}{2}}\{i\lambda(r, a)\}\mathrm{P}_n(\cos\theta), \quad . \quad . \quad . \quad . \quad . \quad (19')$$

where $r>a$, and n may have any integral value in radial flow problems (we replace $n+\frac{1}{2}$ by n in cylindrical flow problems). In accordance with an evaluation given in ((3), p. 238), the corresponding "instantaneous integral" solution is given by

$$v_i = \frac{1}{(ar)^{\frac{1}{2}}.\,2t}e^{-\frac{r^2+a^2}{4\kappa t}}\mathrm{I}_{n+\frac{1}{2}}(\alpha)\mathrm{P}_n(\cos\theta) : \alpha = \left(\frac{ra}{2\kappa t}\right). \quad . \quad . \quad (20)$$

The integration with respect to t' which follows this then yields (as indicated in (5), p. 934),

$$v_c = \frac{1}{(ar)^{\frac{1}{2}}}\cdot\frac{1}{(2n+1)}\cdot\left(\frac{a}{r}\right)^{n+\frac{1}{2}}\mathrm{P}_n(\cos\theta), \quad (r>a), \quad . \quad . \quad . \quad (21)$$

as the corresponding continuous flow solution.

This result may be obtained, however, by a direct application of the theorem in the form expressed by (B) above. If, in (19), we use the well-known expansions for the Bessel functions, the only term in the product which contributes to the continuous flow solution is that of the form $ce^{ikt-i\lambda r}$, in which coefficient c does not contain λ. This term is similar to (12) above, and yields a continuous flow solution as given in (14), in agreement with (21).

It should be noted also that any derivative of v_i, as given by (20), with respect to t, yields zero value for the integration with respect to t', and therefore gives a zero contribution to the steady flow solution.

§ 2. *Sources near a Boundary.*

Point Sources.

(1) *Source within a Spherical Boundary.*

Periodic Solutions.—Let a point source of strength qe^{ikt} be situated at the point $(a', 0)$ within a spherical surface of radius a $(a'<a)$. The complete periodic solution, representing temperature at any point within the surface, is composed of the incident wave

$$v_1=\frac{q}{2\pi\mathrm{K}}\cdot\frac{e^{ikt}}{(ar)^{\frac{1}{2}}}\sum_0^\infty(n+\tfrac{1}{2})p_{n+\frac{1}{2}}\{i\lambda(r, a')\}\mathrm{P}_n(\cos\theta), \quad . \quad . \quad (22)$$

and the reflected wave

$$v_1'=\frac{q}{2\pi\mathrm{K}}\frac{e^{ikt}}{(ar)^{\frac{1}{2}}}\sum_0^\infty(n+\tfrac{1}{2}).\mathrm{A}_n\mathrm{I}_{n+\frac{1}{2}}(i\lambda r)\mathrm{P}_n(\cos\theta) \quad . \quad . \quad . \quad (23)$$

(as given in (3), p. 240, where the coefficients A_n are to be determined from special conditions to be fulfilled at the spherical boundary $r=a$. The complete solution is then given by $v=v_1+v_1'$. The $p_{n+\frac{1}{2}}$ function is defined in (5), p. 925, and in (19) above.

Taking the boundary condition $v=0$, at $r=a$, we find that

$$\mathrm{A}_n=-\frac{\mathrm{K}(z)}{\mathrm{I}(z)}.\mathrm{I}(z'), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

where $z=(i\lambda a)$, $z'=(i\lambda a')$, and K and I represent the Bessel functions of order $(n+\frac{1}{2})$.

Continuous Flow Solutions.

Every term of the incident wave is similar to (19) above, and yields a continuous flow solution similar to (21).

The simplest method of determining the continuous flow solutions corresponding to the reflected waves is as follows:—

Expand the Bessel function ratios, which appear in (24), in a series of ascending powers of λ^2. Thus, in (24), we write

$$\frac{\mathrm{I}(z')}{\mathrm{I}(z)}=\left(\frac{a'}{a}\right)^{n+\frac{1}{2}}+\alpha_1\lambda^2+\alpha_2\lambda^4+,\text{ etc.}, \quad . \quad . \quad . \quad . \quad . \quad (25)$$

where the coefficients α_1, α_2, etc. are functions of a and a', and use this expansion in each term of (23). Then, since λ^2, as in (B) above, may be treated as an operator equivalent to d/dt, it follows in accordance with the theorem given above that the only term in each expansion which contributes to the continuous flow solution is the first term which is independent of λ^2, and of the type discussed in (19).

For the continuous flow problem indicated by the conditions :

(a) Boundary, $r=a$, kept at zero temperature : steady source of strength q at point $(a', 0)$, where $a>a'$, the above method leads to the solution :—

For the incident wave issuing from the source

$$v_1=\frac{q}{4\pi K}\cdot\frac{1}{(ra')^{\frac{1}{2}}}\sum_0^\infty\left(\frac{a'}{r}\right)^{n+\frac{1}{2}}P_n(\cos\theta) \tag{26}$$

$$=\frac{q}{4\pi K r_1}, \tag{26'}$$

where r_1 is the radius vector from the source at $(a', 0)$ to any point within the sphere defined by the co-ordinates (r, θ).

For the reflected wave-system

$$v_1'=-\frac{q}{4\pi K}\cdot\frac{1}{(ra')^{\frac{1}{2}}}\sum_0^\infty\left(\frac{a'}{a}\right)^{n+\frac{1}{2}}\left(\frac{r}{a}\right)^{n+\frac{1}{2}}P_n(\cos\theta). \tag{27}$$

This reflected system may, however, be expressed in terms of a source of strength q', situated outside the sphere at the inverse point with respect to the point $(a', 0)$. By using the relations

$$R=\left(\frac{a^2}{a'}\right) \quad\text{and}\quad \frac{q}{(a')^{\frac{1}{2}}}=\frac{q'}{(R)^{\frac{1}{2}}}, \tag{28}$$

which define R and q', we obtain

$$v_1'=-\frac{q'}{4\pi K R}\sum_0^\infty\left(\frac{r}{R}\right)^n P_n(\cos\theta) \tag{29}$$

$$=-\frac{q'}{4\pi K r_2}, \tag{29'}$$

where r_2 is the radius vector drawn from the image source at $(R, 0)$ to any point within the sphere.

(2) *Source outside a Spherical Boundary.*

Periodic Solutions.

Let the point source of strength qe^{ikt} be situated at point $(R, 0)$, $R>a$, where the origin is the centre of the spherical surface $r=a$. The incident and reflected waves are in this case given by

$$v_1=\frac{q}{2\pi K}\frac{1}{(Rr)^{\frac{1}{2}}}\sum_0^\infty(n+\tfrac{1}{2})p_{n+\frac{1}{2}}\{i\lambda(R, r)\}P_n(\cos\theta), \tag{30}$$

$$v_1' = \frac{q}{2\pi K} \frac{1}{(Rr)^{\frac{1}{2}}} \sum_0^\infty (n+\tfrac{1}{2}) A_n K_{n+\frac{1}{2}}(i\lambda r) P_n (\cos\theta), \quad . \quad . \quad . \quad (31)$$

respectively, in the notation of ⁽³⁾ and ⁽⁵⁾.

The coefficients A_n are to be determined by conditions maintaining at the surface $r=a$. Writing Z to represent $(i\lambda R)$, and taking the condition $v=0$, at $r=a$, we find that

$$A_n = -\frac{K(Z)}{K(z)} I(z), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

K and I being Bessel functions of order $(n+\frac{1}{2})$. When the boundary condition at $r=a$ is $\frac{dv}{dr}=0$, the surface being impermeable to heat, we have

$$A_n = -\frac{zI'(z)-\frac{1}{2}I(z)}{zK'(z)-\frac{1}{2}K(z)} K(Z). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

Equations (30) to (32), referring to an external source, correspond with equations (22) to (24) respectively, referring to an internal source.

Continuous Flow Solutions.

In order to pass from the periodic solutions to the corresponding continuous flow solutions the procedure used in this case is exactly similar to that employed in preceding problems. The method depends on the possibility of expanding the ratio of two Bessel functions of the type $K_n(z)$ in a series of ascending powers of λ^2. For the present we assume the validity of the expansions, which are obtained by using the expression for the $K_n(z)$ Bessel function

$$K_n(z) = \frac{\Gamma(n+\frac{1}{2})}{\sqrt{\pi}} \left(\frac{2}{z}\right)^n \int_0^\infty \frac{\cos(z\sinh\phi)}{\cosh^{2n}\phi} d\phi. \quad . \quad . \quad . \quad . \quad (34)$$

For insertion in (32) the expansion required is

$$\frac{K_{n+\frac{1}{2}}(Z)}{K_{n+\frac{1}{2}}(z)} = \left(\frac{a}{R}\right)^{n+\frac{1}{2}} + \alpha_1\lambda^2 + \alpha_2\lambda^4 +, \text{ etc.}, \quad . \quad . \quad . \quad . \quad (35)$$

and that required in (33) is

$$\frac{K_{n+\frac{1}{2}}(Z)}{zK'_{n+\frac{1}{2}}(z)-\frac{1}{2}K_{n+\frac{1}{2}}(z)} = -\frac{1}{(n+1)} \left(\frac{a}{R}\right)^{n+\frac{1}{2}} + \beta_1\lambda^2 + \beta_2\lambda^4 +, \text{ etc.} \quad . \quad (36)$$

The coefficients α_1, α_2, etc., β_1, β_2, etc., depend on a and R only. The periodic solution is now expressible in terms of functions of the type (19) or their derivatives with respect to t. The continuous solution can therefore be obtained by an application of former results of type (21). Thus for the continuous flow problem indicated by :

(*a*) Boundary $r=a$ at zero-temperature: steady source of heat of strength q at point (R, 0), $R>a$, the following solution is obtained :—

Incident wave issuing from source at (R, 0)

$$v_1=\frac{q}{4\pi K}\,\frac{1}{(Rr)^{\frac{1}{2}}}\sum_0^\infty\left(\frac{r}{R}\right)^{n+\frac{1}{2}}P_n(\cos\theta)\quad . \ . \ . \ . \ . \ . \quad (37)$$

$$=\frac{q}{4\pi K r_1},\quad . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \quad (37')$$

where r_1 is the radius vector drawn to any point external to the sphere from the source at (R, 0).

Reflected system

$$v_1'=-\frac{q}{4\pi K}\,\frac{1}{(rR)^{\frac{1}{2}}}\sum_0^\infty\left(\frac{a}{R}\right)^{n+\frac{1}{2}}\left(\frac{a}{r}\right)^{n+\frac{1}{2}}P_n(\cos\theta)\quad . \ . \ . \quad (38)$$

$$=-\frac{q'}{4\pi K r_2}\ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \quad (38')$$

The latter form is obtained by using the relations in (28) above; r_2 being the radius vector drawn from the inverse point (a', 0) to any point external to the sphere.

These results are again in complete agreement with the image solution of the analogous problem in electrostatics. The image system represented by (26) and (27′) is in fact identical with the image system represented by (37′) and (38′).

For the continuous flow problem indicated by :

(*b*) Boundary $r=a$ impervious to heat: steady source of strength q at external point (R, 0), the solution obtained is :—

Incident wave v_1—as in (37) or (37′).

Reflected system

$$v_1'=\frac{q}{4\pi K}\cdot\frac{1}{(Rr)^{\frac{1}{2}}}\sum_0^\infty\frac{n}{n+1}\left(\frac{a}{R}\right)^{n+\frac{1}{2}}\left(\frac{a}{r}\right)^{n+\frac{1}{2}}P_n(\cos\theta).\quad . \ . \ . \quad (39)$$

When the relations stated in (28) are employed we obtain

$$v_1'=\frac{q'}{4\pi K}\,\frac{1}{(a'r)^{\frac{1}{2}}}\sum_0^\infty\frac{n}{n+1}\left(\frac{a'}{r}\right)^{n+\frac{1}{2}}P_n(\cos\theta),\quad . \ . \ . \quad (39')$$

an equation which represents an image source of strength q' placed at the inverse point (a', 0), together with a line sink extending from the point (a', 0) to the origin—the sink being of uniform strength q' per unit length of the line.

Problems involving steady heat flow near a boundary impervious to heat have analogues in the steady flow of incompressible fluid. The solution given above is in complete agreement with a well known hydrodynamical result.

Line Sources.

If, in the above problems relating to the spherical surface $r=a$, we replace the spherical surface by a cylindrical surface $\rho=a$, at the same time replacing point sources by line sources, we obtain a group of similar problems which can be solved by exactly the same methods. The solutions correspond each to each, and these are stated here in the same order as in the treatment of point source problems. The boundary conditions $v=0$, at $\rho=a$, and $\frac{dv}{d\rho}=0$, at $\rho=a$, are referred to as condition (a) and condition (b) respectively.

Periodic Solutions.

Line source of strength qe^{ikt} per unit length at point $(a', 0)$, $(a>a')$: condition (a).

Incident wave issuing from source

$$v_1=\frac{q}{\pi K}e^{ikt}\sum_0^\infty K_n(i\lambda\rho)I_n(i\lambda a')\cos n\theta, \quad . \quad . \quad . \quad . \quad . \quad (40)$$

in which summation the factor $\frac{1}{2}$ is to be applied to the term corresponding to $n=0$ ([2], p. 705).

Reflected wave

$$v_1'=\frac{qe^{ikt}}{\pi K}\sum_0^\infty A_nI_n(i\lambda\rho)\cos n\theta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

where A_n is to be determined by the boundary condition to be applied at $\rho=a$. For (a)

$$A_n=-\frac{I_n(i\lambda a')}{I_n(i\lambda a)}K_n(i\lambda a). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

Continuous Flow Solutions.

These are derived from the above periodic solutions by the method of this communication.

(a) Incident wave

$$v_1=\frac{q}{\pi K}\sum_0^\infty\frac{1}{2n}\left(\frac{a'}{\rho}\right)^n\cos n\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

$$=-\frac{q}{2\pi K}\log\rho_1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43')$$

where ρ_1 is the radius vector from the source at $(a', 0)$ to any point within the cylinder.

Reflected system

$$v_1'=-\frac{q}{\pi K}\sum_0^\infty\frac{1}{2n}\left(\frac{a'}{a}\right)^n\left(\frac{\rho}{a}\right)^n\cos n\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

$$=+\frac{q}{2\pi K}\log\rho_2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44')$$

where ρ_2 is the radius vector drawn from (R, 0), the inverse point to $(a', 0)$, to any point within the cylinder. R is again defined by (28) above.

Periodic Solutions.

Line source of strength qe^{ikt} per unit length at point (R, 0), $R>a$: conditions (a) and (b).

Incident wave issuing from the source

$$v_1=\frac{q}{\pi K}e^{ikt}\sum_0^\infty K_n(i\lambda R)I_n(i\lambda\rho)\cos n\theta. \quad (R>\rho) \qquad (45)$$

Reflected wave

$$v_1'=\frac{q}{\pi K}\,.\,e^{ikt}\sum_0^\infty A_nK_n(i\lambda\rho)\cos n\theta. \qquad (46)$$

For (a),
$$A_n=-\frac{K_n(i\lambda R)}{K_n(i\lambda a)}I_n(i\lambda a). \qquad (47)$$

For (b),
$$A_n=-\frac{K_n(i\lambda R)}{K_n'(i\lambda a)}I_n'(i\lambda a). \qquad (48)$$

Corresponding Continuous Flow Solutions.

For (a). Incident wave

$$v_1=\frac{q}{\pi K}\sum_0^\infty\frac{1}{2n}\left(\frac{\rho}{R}\right)^n\cos n\theta \qquad (49)$$

$$=-\frac{q}{2\pi K}\log\rho_2, \qquad (49')$$

where ρ_2 is radius vector drawn from the source at (R, 0) to any point external to the cylinder.

Reflected system

$$v_1'=-\frac{q}{\pi K}\sum_0^\infty\frac{1}{2n}\left(\frac{a}{R}\right)^n\left(\frac{a}{\rho}\right)^n\cos n\theta \qquad (50)$$

$$=+\frac{q}{2\pi K}\log\rho_1, \qquad (50')$$

where ρ_1 is radius vector from the inverse point $(a', 0)$.

For (b). Incident wave—v_1 as in (49).

Reflected system

$$v_1'=\frac{q}{\pi K}\sum_1^\infty\frac{1}{2n}\left(\frac{a}{R}\right)^n\left(\frac{a}{\rho}\right)^n\cos n\theta. \qquad (51)$$

No term appears corresponding to $n=0$. Hence, making use of (28), we can write

$$v_1'=-\frac{q}{2\pi K}(\log\rho_1-\log\rho), \qquad (51')$$

where ρ_1 is the radius vector drawn from the inverse point $(a', 0)$ to any point of the medium, and ρ is the radius vector from the origin to the same point.

In all the problems considered above relating to sources near a boundary which is kept at zero temperature, the continuous flow solutions can be interpreted in terms of sources and their images in the boundary sphere or cylinder. Each problem has its electrostatic analogue, with which results obtained above are in complete agreement. The image solution represented by (26′) and (29′) is obtained by means of the expansion, shown in (25), of the ratio of two Bessel functions of type I_n. The same image solution is represented by (37′) and (38′), but is obtained by means of the expansion shown in (35) of the ratio of two Bessel functions of type K_n.

Similarly, the continuous flow solutions obtained above, relating to a boundary which is impervious to heat, apply equally to flow of an incompressible fluid. Each problem has its hydrodynamical analogue, with which results obtained above are in complete agreement.

Difficulties arise in certain problems in the application of the method used in this investigation. The above results, however, serve to test and to confirm the validity of the method in general.

List of References.

(1) Phil. Mag. iii. pp. 784–800 (April 1927, Suppl.).
(2) *Ibid.* v. pp. 701–720 (April 1928).
(3) *Ibid.* xii. pp. 233–255 (August 1931, Suppl.).
(4) *Ibid.* ser. 7, xviii. pp. 625–640 (Oct. 1934).
(5) *Ibid.* ser. 7, xxi. pp. 922–934 (May 1936).
(6) *Ibid.* ser. 7, xxi. pp. 934–947 (May 1936).
(7) *Ibid.* ser. 7, xxii. pp. 1079–1088 (Dec. 1936).
(8) *Ibid.* ser. 7, xxix. pp. 100–102 (Jan. 1940).

XIV. *Approximate Formulæ for Functions expressed as Definite Integrals.*

By A. Majid Mian, Ph.D.*, and S. Chapman, F.R.S.†

[Received November 10, 1941.]

Summary.

In the theory of absorption of radiation in an exponential atmosphere it is necessary to consider the integrals

$$E_n(t)=\int_1^\infty e^{-tx}x^{-n}\,dx,\quad B_n(t)=\frac{2}{\pi}\int_0^{\frac{1}{2}\pi} e^{-t\sec\phi}\cos^n\phi\,d\phi.$$

The available series-expressions for these integrals are inconvenient for numerical calculation for moderate (integer) values of n, and for values of t that are neither small nor large (say from 1 to 5). Approximate expressions $E'_n(t)$ and $B'_n(t)$ are therefore given here, in the form of finite series (or *index-sums*)

$$E'_n(t)=\Sigma(e_m/\epsilon_m^{n-2})e^{-\epsilon_m t},\quad B'_n(t)=\Sigma(b_m/\beta_m^n)e^{-\beta_m t};$$

four or five terms suffice to give a good approximation over a considerable range of n and t.

Such index-sum approximations are convenient not only for numerical calculation, but also for use when the integrals $E_n(t)$ or $B_n(t)$ occur as factors in the integrands of further definite integrals. They also lend themselves to the calculation of various linear combinations of E_n (or B_n) for different values of n.

In § 6 it is shown that such index-sums may be regarded as approximations to the sums of the partial integrals obtained by subdividing the range of integration of E_n or B_n.

---

1. The integrals

$$\int_1^\infty e^{-tx}x^{-n}\,dx,$$

where n is zero or a positive integer, are of importance in the theory of absorption of radiation in an atmosphere whose density varies exponentially with the height. Various notations for the integral have been used by writers on this subject: E. Gold (1909) used $H_n(t)$, and was followed in this usage by O. F. T. Roberts (1930); E. Hopf (1934) used $E_n(t)$; A. Unsöld (1938) used $K_n(t)$, and A. Hammad and S. Chapman (1939) used $Ej_n(t)$. Here we shall follow Hopf in using the notation $E_n(t)$. These writers, as also Jahnke and Emde (in their 'Funktionentafeln'), and Nielsen (1906), have given asymptotic and recurrence formulæ for $E_n(t)$, and other information concerning it.

---

* Part of this paper was included in a Ph.D. thesis of the University of London.

† Communicated by the Authors.

Of these we may quote the following:

$$E_n(0)=1/(n-1) \quad (n>1). \tag{1}$$

$$dE_n(t)/dt=-E_{n-1}(t). \tag{2}$$

$$(n-1)E_n(t)=e^{-t}-tE_{n-1}(t) \quad (n>1). \tag{3}$$

$$E_n(t)=e^{-t}\left\{\frac{1}{n-1}-\frac{t}{(n-1)(n-2)}+\ldots+\frac{(-1)^{n-2}t^{n-2}}{(n-1)!}\right\}+\frac{(-1)^{n-1}t^{n-1}}{(n-1)!}E_1(t). \tag{4}$$

$$E_n(t)\sim e^{-t}\left\{\frac{1}{t}-\frac{n}{t^2}+\frac{n(n+1)}{t^3}-\ldots\right\}. \tag{5}$$

$$\text{Lim}\,\{E_1(t)+\log t\}=-\gamma \quad \text{as } t\to 0. \tag{6}$$

$$-E_1(t)=\gamma+\log t-t+\sum_{n=2}^{\infty}(-1)^n t^n/n!\,(n+1). \tag{7}$$

$$E_2(t)=t\log t+1+(\gamma-1)t-\sum_{n=2}^{\infty}(-1)^n t^n/n!\,(n+1). \tag{8}$$

2. The integral $E_1(t)$ is the same as that which Jahnke and Emde, and also J. B. Dale, denote by $-E_1(-t)$, and call the exponential integral Many tables of this function have been published by Glaisher, J. B. Dale, Jahnke and Emde and others. Gold gave a table of E_1, E_2 and E_3 from $t=0{\cdot}01$ to $t=6$.

From the tables of E_1 it is possible, by means of (3), to calculate values of E_n for $n>1$; but for values of t of several units or more, the errors in the last decimal place quickly increase, if 5-figure tables of E_1 are used.

One may calculate $E_n(t)$ by formulæ such as (7, 8) for small values of t, and from (4) for large values. But neither type of formula is convenient for intermediate values, and in this range (say from 1 to 5) approximate formulæ are desirable. Roberts gave the crude approximation $\frac{1}{2}e^{-1{\cdot}5t}$ for $E_3(t)$; this is only very roughly satisfactory, as is shown by the following table:—

$t=$	0	0·01	0·05	0·1	0·2	0·5	1·0	1·5	2·0	5·0
$1000\,E_3(t)=$	500	490	455	416	352	222	110	57	30	1
$500\,e^{-1{\cdot}5t}=$	500	493	464	430	370	236	112	53	25	0

3. We have had occasion to devise better (and necessarily more complicated) approximate formulæ for $E_n(t)$, for integer values of n from 2 onwards, and for moderate values of t. It is convenient to denote our approximate expression by $E'_n(t)$, to distinguish it from $E_n(t)$. We take

$$E'_n(t)=\sum_m e_{m,n}e^{-\epsilon_{m,n}t}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where $e_{m,n}$ and $\epsilon_{m,n}$ are positive : it is convenient to order the terms so that $\epsilon_{m,n}$ increases with m. Each term of the series (1) is positive, and decreases steadily to zero as $t\to 0$. The number of terms we take to be finite : a good approximation over a considerable range of n and t can be obtained with four or five terms.

It is found to be possible to take values of $\epsilon_{n,m}$ that are independent of n, so that we drop the suffix n ; we also take

$$e_{m,n}=e_{m,2}/\epsilon_m^{n-2}\equiv e_m/\epsilon_m^{n-2}, \quad n\geqslant 2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

(replacing $e_{m,2}$ by e_m), so that, analogously to § 1(2),

$$dE'_n(t)/dt=-E'_{n-1}(t), \quad n\geqslant 3. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Consequently we may replace (1) by

$$E'_n(t)=\sum_m (e_m/\epsilon_m^{n-2})e^{-\epsilon_m t}, \quad n\geqslant 2 \; ; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

thus, if five terms are taken, E'_n involves five values of e_m and five of ϵ_m.

It is convenient to call (4) an *index-sum*, and e_m, ϵ_m its (index) coefficients and (index) factors respectively.

For numerical calculation one may re-write (4) in the convenient form

$$E'_n(t)=\sum_m 10^{-\epsilon'_m t+\log_{10}e_m-(n-2)\log_{10}\epsilon_m}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where $\epsilon'_m=\epsilon_m\log_{10}e$; knowing ϵ'_m and e_m, it is easy to calculate the total index in each term, and then by looking at a table of antilogarithms to evaluate the term.

At the outset we choose to impose the condition that

$$E'_2(0)=E_2(0)=1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

so that

$$\sum e_m=1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

4. The index-coefficients and factors, e_m and ϵ_m, may be determined in various ways ; one convenient method is to determine them by making $E'_n(0)$ agree with the known values of $E_n(0)$ for a series of values of n (*cf.* § 14). By this method we have obtained the following values :—

$m=$	1	2	3	4	5
$e_m=$	0·09846	0·24048	0·32212	0·24048	0·09846
$\epsilon_m=$	1·03595	1·2610	2	4·8310	26·316

5. Using these values, and expressing $E'_n(t)$ in the form § 3(5), we have calculated values of $10^5E'_n(t)$ for $n=2$ to $n=8$, and from $t=0{\cdot}01$ to $t=6$; in the following tables these are compared with the values of $10^5E_n(t)$ based on Dale's 5-figure tables of $E_1(t)$. After each value of $10^5E_n(t)$ is given (in brackets) the value of $10^5\Delta E'_n$, defined by

$$\Delta E'_n \equiv E'_n(t)-E_n(t).$$

Except for $E_2(t)$ for $t<0{\cdot}25$, the differences are remarkably small; even for $E_2(t)$ the greatest error is only 0·6 per cent.; the small errors in the rest of the table may be due partly to slight errors in the calculated values of $E_n(t)$ as well as in those of $E'_n(t)$.

For the larger values of t in the following table, where $E_n(t)$ becomes small, the comparison scarcely indicates how close is the approximation of $E'_n(t)$ to $E_n(t)$. This can be shown better by evaluating $e^t\Delta E'_n$, since e^tE_n and $e^tE'_n$ decrease much less rapidly, as t increases, than E_n or E'_n.

For the values of t in the table, from $t=0{\cdot}5$ to $t=6$, we find that for $n=2$ the values of $10^4e^t\Delta E'_n$ are respectively 0, 1, -3, -1, -3, -4, -1, -6, -5; for the integer values of t from 7 to 14 they are 4, 0, 1, -4, -6, -1, -1, 0. Similarly, for $n=3$, and for $t=0{\cdot}5$ to $t=6$, they are 0, 1, 1, -1, -2, 3, 6, 5, 2.

Thus, except for a small initial range of t in the case of $E_2(t)$, the index-sum $E'_n(t)$ provides a good approximation to $E_n(t)$, accurate enough for most numerical calculations in atmospheric problems, up to $t=10$ or more, and at least up to $n=8$.

It would doubtless be possible to modify our approximate formula for $E_n(t)$ so as to improve its accuracy, either over the range of n and t considered in the above table, or for E_2 for small values of t. But for the purpose of most atmospheric problems, at least, the approximation seems sufficient already.

An alternative four-term index-sum $E'_n(t)$, with different numerical values of e_m and ϵ_m, is given in § 15.

6. Some insight into the reason for the good approximation given by $E'_n(t)$ to $E_n(t)$ may be gained as follows:—

In $E_n(t)$, change the variable from x to $y(=1/x)$, so that the range of integration is made finite (namely 1); thus

$$E_n(t)=\int_0^1 e^{-t/y}y^{n-2}\,dy. \quad \ldots\ldots\ldots \quad (1)$$

For $n\geqslant 2$ the integrand in (1), as well as the range, is finite.

If we divide the range of integration in (1) into a number of finite parts, 0 to e_1, e_1 to e_1+e_2, and so on, so that $\Sigma e_m=1$, the partial integrals so obtained may be typified by

$$\int_{y_m}^{y_{m+1}} f(n,t,y)\,dy, \quad \ldots\ldots\ldots \quad (2)$$

TABLE.

Values of $10^5 E_n(t)$ and, in brackets, of $10^5 \Delta E'_n(t)$.

t.	$10^5 E_2(t)$.	$10^5 E_3(t)$.	$10^5 E_4(t)$.	$10^5 E_5(t)$.	$10^5 E_6(t)$.	$10^5 E_7(t)$.	$10^5 E_8(t)$.
0	100000 (0)	50000 (−1)	33333 (−2)	25000 (−4)	20000 (−2)	16667 (0)	14286 (1)
0·01	94967 (579)	49027 (−6)	32838 (−2)	24669 (−3)	19752 (−2)	16468 (0)	14120 (1)
0·05	82783 (−180)	45492 (−14)	30949 (−2)	23394 (−3)	18791 (−2)	15697 (1)	13477 (1)
0·1	72255 (−269)	41629 (1)	28773 (−1)	21902 (−3)	17659 (−2)	14787 (1)	12715 (1)
0·25	51773 (103)	32468 (6)	23254 (−2)	18016 (−2)	14675 (−1)	12369 (0)	10684 (1)
0·5	32665 (−10)	22160 (−7)	16524 (−2)	13098 (−2)	10821 (−1)	9207 (1)	8007 (1)
1·0	14850 (−2)	10969 (3)	8606 (−1)	7046 (−2)	5948 (1)	5040 (1)	4521 (0)
1·5	7310 (9)	5674 (0)	4601 (−2)	3853 (0)	3307 (1)	2892 (1)	2568 (−1)
2·0	3754 (0)	3013 (−2)	2503 (−2)	2133 (1)	1854 (1)	1638 (0)	1465 (0)
2·5	1982 (−5)	1627 (1)	1379 (−1)	1190 (1)	1047 (0)	932 (0)	840 (−1)
3·0	1064 (−2)	894 (−1)	766 (1)	670 (0)	594 (0)	533 (0)	483 (−1)
4·0	320 (0)	276 (0)	243 (0)	216 (−1)	194 (0)	176 (0)	161 (0)
5·0	99 (0)	90 (−2)	75 (3)	75 (−4)	60 (4)	62 (−3)	52 (2)
6·0	32 (0)	28 (0)	27 (−1)	22 (1)	23 (−2)	18 (2)	20 (−2)

where

$$y_{m+1}-y_m=e_m,\quad y_0=0,\quad y_m=e_1+e_2+\ldots+e_{m-1},\quad f(n,t,y)=e^{\ t/y}y^{n-2}.$$

By the mean value theorem for integrals, (2) is equal to

$$e_m f(n,t,\mathrm{Y}_{m,n,t}), \quad \ldots \ldots \quad (3)$$

where $\mathrm{Y}_{m,n,t}$ is some value of y between y_m and y_{m+1}. As the suffixes indicate, this value will in general depend on m, n and t. Hence, adding the values of the partial integrals, so expressed, and writing $f(n,t,\mathrm{Y}_{m,n,t})$ in full, we have

$$\mathrm{E}_n(t)=\Sigma e_m\mathrm{Y}_{m,n,t}^{n-2}e^{-t/\mathrm{Y}_{m,n,t}}. \quad \ldots \ldots \quad (4)$$

If the ranges e_m are not too large, the values $\mathrm{Y}_{m,n,t}$ will lie not far from the middle of the corresponding intervals y_m to y_{m+1}, and for a moderate range of values of n and t their positions in their respective intervals will not vary much with n and t. Hence one may expect that it will be possible to choose *fixed* values of $\mathrm{Y}_{m,n,t}$, independent of n and t, and therefore conveniently denoted by Y_m, such that the *sum*

$$\Sigma e_m\mathrm{Y}_m^{n-2}e^{-t/\mathrm{Y}_m} \quad \ldots \ldots \quad (5)$$

will approximate to (4), and therefore to $\mathrm{E}_n(t)$, over moderate ranges of n and t, even though the individual terms in (4) and (5) are not respectively equal to one another, and may for some values of n and t differ from one another by more than their sums do. This expectation is confirmed in § 5, because by writing $\mathrm{Y}_m=1/\epsilon_m$, (5) becomes identical with the index-sum § 3 (4) there examined.

It is easy to verify that the reciprocals of the numerical values of ϵ_m given in § 4 lie near the middle of the intervals y_m to y_{m+1}, corresponding to the values of e_m given in § 4.

If five terms are taken in the index-sum, ten disposable constants are involved, or nine if we impose the condition § 3 (7). It is not surprising that with nine disposable constants it is possible to approximate by (5) to the function $\mathrm{E}_n(t)$, of two variables n and t, over a considerable range of n and t.

The method of determining the nine constants, by making $\mathrm{E}'_n(0)$ approximate to $\mathrm{E}_n(0)$ for several values of n, ensures that $\mathrm{E}_n(t)-\mathrm{E}'_n(t)$ shall be small for $t=0$ and for small values of t. But in the absence of a proper theory of the error of the approximation, as a function of n and t, it is impossible to be sure beforehand that this method of determining e_m and ϵ_m will give a satisfactory approximation to $\mathrm{E}_n(t)$ for values of t that are not small. For the present the method is a purely empirical one, though it gives results that are likely to be helpful and satisfactory to the applied mathematician.

7. The discussion of atmospheric absorption leads also to the consideration of the integrals

$$\mathrm{B}_n(t)=\frac{2}{\pi}\int_0^{\frac{1}{2}\pi} e^{-t\sec\phi}\cos^n\phi\,d\phi, \quad \text{. . . . .} \quad (1)$$

and we have been led to devise approximate formulæ, in the form of index-sums, for these integrals. Our approximate expressions will be denoted by $\mathrm{B}'_n(t)$ or $\mathrm{B}''_n(t)$.

The following properties of $\mathrm{B}_n(t)$ are readily proved :—

$$\mathrm{B}_{2n}(0)=\frac{(2n-1)\,!}{2^{2n}(n\,!)^2}, \quad \mathrm{B}_{2n+1}(0)=\frac{2^{2n+1}(n\,!)^2}{(2n+1)\,!\,\pi}. \quad \text{. . . . .} \quad (2)$$

$$\partial\mathrm{B}_n(t)/\partial t=-\mathrm{B}_{n-1}(t). \quad \text{. . . . .} \quad (3)$$

$$\mathrm{B}_n(t)\sim\left(\frac{2}{\pi t}\right)^{1/2} e^{-t}\left\{1-\frac{4n+5}{2^3t}+\frac{3(16n^2+56n+43)}{2.2^6t^2}\dots\right\}. \quad \text{. . .} \quad (4)$$

$$\mathrm{B}_0(t)\sim\left(\frac{2}{\pi t}\right)^{1/2} e^{-t}\left\{1-\frac{5}{8t}+\frac{1{\cdot}0079}{t^2}-\frac{2{\cdot}5928}{t^3}+\frac{9{\cdot}1869}{t^4}\dots\right\}. \quad \text{. .} \quad (5)$$

$$\mathrm{B}_0(t)=\frac{2}{\pi}\int_t^\infty \mathrm{K}_0(x)\,dx \quad \text{. . . . .} \quad (6)$$

where $\mathrm{K}_0(x)$ is the well-known Bessel function : *cf.* Watson's 'Bessel Functions,' pp. 80, 77, 60), so that

$$\begin{aligned}\mathrm{B}_0(t)=1&-\frac{2}{\pi}\left\{\log\left(\frac{2}{t}\right)-\gamma\right\}\left\{t+\frac{t^3}{2^2.3}+\frac{t^5}{2^2.4^2.5}+\dots\right\}\\ &-\frac{2}{\pi}\left\{\frac{t^3}{2^3.3}+\frac{t^5}{2^2.4^2.5}(1+\tfrac{1}{2})+\frac{t^7}{2^2.4^2.6^2.7}(1+\tfrac{1}{2}+\tfrac{1}{3})+\dots\right\}\\ &-\frac{2}{\pi}\left\{t+\frac{t^3}{2^2.3^2}+\frac{t^5}{2^2.4^2.5^2}+\dots\right\}. \quad \text{. . . . .} \quad (7)\end{aligned}$$

$$\mathrm{B}_n(t)=(n-1)\mathrm{B}_{n-2}(t)+t\{\mathrm{B}_{n-3}(t)-\mathrm{B}_{n-1}(t)\}. \quad \text{. . . . .} \quad (8)$$

In the last equation, when $n<3$, $\mathrm{B}_{-1}(t)$ must be replaced by $(2/\pi)\mathrm{K}_0(t)$, and $\mathrm{B}_{-2}(t)$ by $(2/\pi)\mathrm{K}_1(t)$; both these functions are tabulated in Watson's 'Bessel Functions.' From (7) and (8) it is therefore possible to calculate successively B_1, B_2, B_3, and so on.

Except when t is small, (7) and (8) are not quickly convergent, and for $1<t<5$ their use for the calculation of B_0 and, still more, of B_n, is very laborious. The series (5) (to t^{-4}) gives $\mathrm{B}_0(t)$ within a few per cent. when $t=5$, and within 1 per cent. when $t>8$. When $n>0$, more terms of the asymptotic series must be taken to give the same accuracy. For moderate values of t these asymptotic formulæ for $\mathrm{B}_n(t)$ are of little practical use.

8. We have found it possible to represent $B_n(t)$ over a considerable range of n and t by an index-sum of five terms, as was done for $E_n(t)$ in § 3. Thus, corresponding to (4) and (7) of § 3, let

$$B'_n(t)\equiv\sum_m(b_m/\beta_m^n)e^{-\beta_m t},\quad \Sigma b_m=1. \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

This approximation can be regarded as a sum of partial integrals, as in § 6, if we write

$$B_n(t)=\int_0^1 e^{-\ \sec\phi}\cos^n\phi\, d(\phi/\tfrac{1}{2}\pi). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

If $\beta_m=\sec\phi_m$, then b_m denotes a range of the variable of integration $\phi/\frac{1}{2}\pi$ in (2), and ϕ denotes a value of ϕ within the corresponding range.

9. We have obtained several alternative index-sums for B'_n, of which two only will be cited it is convenient to denote the second form by B''_n.

First form, B'_n	m	1	2	3	4	5
	b_m	0·20878	0·22386	0·25318	0·24626	0·06792
	β_m	1·0134	1·1402	1·5588	3·4849	59·13
Second form B''_n	m	1	2	3	4	5
	b_m	0·2	0·2	0·2	0·2	0·2
	β_m	1·0123	1·1224	$\sqrt{2}$	2·2022	6·4282

The first form was obtained by choosing the disposable constants so that $B'_0(0)=B_0(0)=1$, and so that $B'_{2n}(0)$ should very nearly equal $B_{2n}(0)$ for nine other values of n. In the second form all the values of b_m are equal, and only two disposable constants had to be found empirically (§ 13). In calculating B''_0 the equality of the b's in B''_0 is advantageous, but this advantage is not preserved for $n>0$.

For $n=t=0$, $\partial B_n(t)/\partial t=\infty$, so that for $n=0$ any index-sum approximation must be deficient near (though not *at*) $t=0$; when $t=0$, $\partial B'_0(t)/\partial t=5{\cdot}756$, and $\partial B''_0(t)/\partial t=2{\cdot}436$, so that near $t=0$ the better approximation to B_0 is given by B'_0. Actually we find that $B''_1(0)$ is in error by about 3 per cent., though for other values of n the error of $B''_n(0)$ is smaller than that of $B'_n(0)$, as is shown by the following table, wherein

$$\Delta B'_n(t)\equiv B'_n(t)-B_n(t)$$

and likewise for $\Delta B''_n(t)$.

$n=$	0	1	2	3	4	5	6	7
$10^4\Delta B'_n(0)$	0	0·2	0	1	0·4	0·3	1·0	1
$10^4\Delta B''_n(0)$	0	−26·2	0	0·4	0	−0·1	0	0

When t is small though not zero, $10^4B_0(t)$ may conveniently be calculated from § 7 (7); in this way, for $t=0{\cdot}04$, 0·1 and 0·5 we obtain the values 8896, 7831 and 4098. The corresponding values of $10^4\Delta B_0'$ and $10^4\Delta B_0''$ are respectively -167, -24, 18 and 205, 157 and -20.

The following table gives $10^4e^tB_n'(t)$ for integer values of t from 1 to 8, for $n=0$ and $n=1$, and (in brackets) the difference $10^4e^t\{B_n''(t)-B_n'(t)\}$; the factor e^t is inserted because otherwise the magnitudes to be compared would be much smaller for the larger values of t.

$t=$	1	2	3	4
$10^4e^tB_0'(t)$	5659 (18)	4575 (-4)	3952 (-6)	3529 (-2)
$10^4e^tB_1'(t)$	4728 (9)	4028 (-6)	3573 (-1)	3247 (2)

$t=$	5	6	7	8
$10^4e^tB_0'(t)$	3218 (4)	2980 (4)	2790 (3)	2635 (1)
$10^4e^tB_1'(t)$	3000 (4)	2804 (4)	2644 (0)	2510 (1)

This table shows that the two index-sums B_n' and B_n'', for $n=0$, $n=1$, are accordant over this range of t within a small fraction of 1 per cent. The exact values of $10^4e^tB_0(t)$ for $t=1$ and $t=2$ have been calculated from § 7 (7): they are 5680 and 4568, so that the corresponding values of $10^4e^t\Delta B_n''(t)$ are -3 and 3: here B_0'' is more exact than B_0'. For $t>2$ the series § 7 (7) is unsuitable for the calculation of B_0.

In the absence of accurate values of $B_0(t)$ and $B_1(t)$ between 2 and 7 we are unable to determine the corresponding values of $\Delta B_n'$ and $\Delta B_n''$; but beyond $t=8$ it is possible to compare our index-sums with the asymptotic formulæ (4, 5) for B_0 and B_1: the errors are found to be only a small fraction of 1 per cent. Hence it seems probable that these approximations are valid within less than 1 per cent., except near $t=0$ for $n=0$ and $n=1$. The close agreement between the two different index-sums B_n' and B_n'' for most values of n and t indicates that there is probably a considerable variety of suitable index-sums for $B_n(t)$; those here given are unlikely to be the best.

Regarding the index-sums as sums of partial integrals, in B_n'' the five ranges of integration of $\phi/\frac{1}{2}\pi$ are equal; in both B_n' and B_n'' the values ϕ_m of ϕ corresponding to β_m lie nearly though not quite at the centres of the corresponding ranges of integration.

10. If we have an index-sum for one integral, such as E_n or B_n, the same index-coefficients and factors can be used in index-sums for associated integrals; interpreting the index-sums as sums of partial integrals, this signifies merely that we take for the new integral the same partial ranges of integration as for the original integral, and also the same particular value of the variable of integration within each range. The procedure can only be regarded as tentative and suggestive; the errors of the resulting index-sum need then to be checked by comparison with any available exact values of the integral. Nevertheless, numerous

examples can be given where the approximations thus found are satisfactory for many practical purposes.

11. Thus, if we represent the integral

$$C_n(t)=\frac{2}{\pi}\int_0^{\frac{1}{2}\pi} e^{-t\sec^2\phi}\cos^{2n}\phi\,d\phi$$

by the approximation

$$C'_n(t)=\sum_m (c_m/\gamma_m^n)e^{-\gamma_m t},$$

we may take $c_m=b_m$, $\gamma_m=\beta_m^2$, on the analogy of the sums of partial integrals (§ 8). Moreover, since

$$C_n(0)=B_{2n}(0),\quad C'_n(0)=B'_{2n}(0),$$

any index-sum for B_{2n} which is satisfactory for all values of $2n$ over a given range of n will thus yield an index-sum $C'_n(t)$, which for $t=0$ will be satisfactory for the same range of n. The index-sum will also, like C_n, satisfy a relation analogous to § 1(2) or § 7(3).

This approximation, for the particular index-sum derived from the $B'_n(t)$ of § 9, can be checked as follows, by means of the error integral:

$$I(z)=\frac{2}{\sqrt{\pi}}\int_z^\infty e^{-x^2}dx$$

$$=\frac{4}{\pi}\int_z^\infty\int_0^\infty e^{-x^2-y^2}dx\,dy=\frac{4}{\pi}\int_0^{\frac{1}{2}\pi}\int_{z\sec\theta}^\infty e^{-r^2}r\,dr\,d\theta$$

$$=\frac{2}{\pi}\int_0^{\frac{1}{2}\pi} e^{-z^2\sec^2\theta}d\theta=C_0(z^2).$$

We will therefore compare $I(z)$ and $C'_0(z^2)$ for large values of z (*i. e.*, $z^2>10$) and for smaller values ($z^2=\frac{1}{4}$ to $z^2=4$); for the large values only the first two terms of C'_0, or at most three, enter into the calculation, which is therefore less laborious than the calculation of $I(z)$ from the known asymptotic series.

$z^2=0$	$\frac{1}{4}$	1	2·25	4
$1-I(z)=0$	0·5205	0·8427	0·9661	0·9953
$1-C'_0(z^2)=0$	0·5276	0·8420	0·9662	0·9955

For the large values of z we compare $\epsilon^{z^2}I(z)$ with $\epsilon^{z^2}C'_0(z^2)$, because I and C'_0 are themselves so very small for these values of z.

$z^2=$ 10	15	20	25
$10^4\epsilon^{z^2}I(z)=1696$	1409	1229	1104
$10^4\epsilon^{z^2}C'_0(z^2)=1705$	1417	1222	1063

evidently for very large values of z^2 the proportionate error of C'_0 must dcrease, though the actual error, like I itself, is very small. These tables show that from $z^2=\frac{1}{4}$ to $z^2=20$ the index-sum provides a close approximation to $I(z)$; it is convenient for the approximate calculation of $I(z)$ over a range of z intermediate between those for which the power series (z small) and asymptotic series (z large) are suitable.

Doubtless our index-sum for C_0', derived from that given in § 7 for B_0', could be improved; even without this, however, it should be useful in calculations that involve integrals in whose integrand $I(z)$ occurs as a factor; the replacement of $I(z)$ by the index-sum may render approximate integration practicable.

12. As another example, for the integral

$$D_n(t)=\frac{2}{\pi}\int_0^{\frac{1}{2}\pi} e^{-t\cos\phi}\cos^n\phi\,d\phi, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

we may take as an approximation the index-sum

$$D_n'(t)=\sum_m (b_m/\beta_m^n)e^{-t\beta_m}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where the numbers b_m and β_m are those introduced in § 8. For numerical comparison, we take in the present case the second set of values there considered, relating to B_n'', which for $n=0$ gives

$$D_0'(t)=\frac{1}{5}(e^{-0·9878t}+e^{-0·89095t}+e^{-0·7071t}+e^{-0·4541t}+e^{-0·1556t}). \quad . \quad (3)$$

For numerical comparison let us take $t=ix$, in which case the real part of $D_0(ix)$ is the Bessel function $J_0(x)$; the corresponding real part of (3) is obtained by replacing the exponential functions by cosines. Taking $x=1$, 2, 3 and 4, we have the following comparison:

$x=$	1	2	3	4
$J_0(x)=$	0·7652	0·2239	0·2601	0·3971
Real part of (3)=	0·7652	0·2224	0·2606	0·3968

If, instead, we take the values of b_m and β_m in B_n' (§ 8), we obtain an alternative cosine series for $J_0(x)$, which when calculated to four decimal places, for $x=0$, $\frac{1}{2}$, 1, 2, 3, 4 and 6, exceeds $J_0(x)$ by $10^{-4}\times(0, 0, -3, 3, 4, 10, -24)$ respectively.

The series thus derived for $J_0(x)$ are a generalization of the ordinary Fourier's or trigonometric series, in which the factors of the arguments are integers.

13. The index-sum B_n'' of § 8, in which only two disposable constants had to be found empirically, we obtained as follows:—

We consider the integrals

$$S_n(t)=\frac{2}{\pi}\int_0^{\frac{1}{2}\pi}=e^{-t\cos^2\phi}\cos^{2n}\phi\,d\phi, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

which we represent approximately by the index-sum

$$\Sigma(b_m\beta_m^n)e^{-\beta_m t}. \qquad (2)$$

Evidently

$$\frac{2}{\pi}\int_0^{\frac{1}{2}\pi} e^{-t\cos^2\phi}\,d\phi=\frac{2}{\pi}\int_0^{\frac{1}{2}\pi} e^{-t\sin^2\phi}\,d\phi=e^{-t}\left\{\frac{2}{\pi}\int_0^{\frac{1}{2}\pi} e^{t\cos^2\phi}\,d\phi\right\}. \qquad (3)$$

By (2), we approximate to the first of the integrals in (3) by the index-sum $\Sigma b_m e^{\beta_m t}$; by reversing the sign of t this gives an index-sum for the last integral on the right of (3); and as an approximation to (3) we therefore have

$$\Sigma b_m e^{-\beta_m t}=e^{-t}\Sigma b_m e^{\beta_m t}=\Sigma b_m e^{-(1-\beta_m)t}. \qquad (4)$$

This suggests the division of the range of integration (with respect to $\phi/\frac{1}{2}\pi$) symmetrically into an odd number of parts such that the first and last are equal, and likewise the second and last but one, and so on; when this is done, by combining the two forms in (3), and halving, we obtain the index-sum

$$S_0'(t)=\tfrac{1}{2}\Sigma b_m(e^{-\beta_m t}+e^{-(1-\beta_m)t}). \qquad (5)$$

In particular, if we take five equal ranges of integration so that $b_m=\frac{1}{5}$, we have

$$S_0'(t)=\frac{1}{10}\sum_{1,2}\{e^{-\beta_m t}+e^{-(1-\beta_m)t}\}+\frac{1}{5}e^{-\frac{1}{2}t}. \qquad (6)$$

Evidently β_m and $1-\beta_m$ will correspond to $\cos^2\phi_m$ and $1-\cos^2\phi_m$, where ϕ_m is a value of ϕ within the corresponding range (of width b_m), as in § 8.

By n-fold differentiation we obtain from (6) the index-sum

$$S_n'(t)=\frac{1}{10}\sum_{1,2}\{\beta_m^n e^{-\beta_m t}+(1-\beta_m)^n e^{-(1-\beta_m)t}\}+\frac{e^{-\frac{1}{2}t}}{(5.2^n)}, \qquad (7)$$

and, in particular,

$$S_n'(0)=\frac{1}{10}\sum_{1,2}\{\beta_m^n+(1-\beta_m)^n\}+\frac{1}{(5.2^n)}. \qquad (8)$$

In (7) and (8) there are only two unknowns, β_1 and β_2. These can be found by taking two values of n in (8), and determining β_1 and β_2 so that (8) is satisfied for $S_n'(0)=S_n(0)$; the latter is known for all values of n. It was in this way that the index-sum $B_n''(t)$ of § 8 was found. The corresponding values of ϕ_m are nearly, but not exactly, equal to 9°, 27°, 45°, 63° and 81°.

14. The index-sum $E_n'(t)$ of § 4 was found by making use of a symmetry property analogous to that indicated in § 13 (3, 4). We consider the known integral

$$F_0(t)=\int_0^1 e^{-tx}\,dx. \qquad (1)$$

This is unaltered by changing the variable of integration from x to $1-x$; hence we proceed as in § 13 by dividing up the range symmetrically (but not now equally) into an odd number of parts, so that the first and last intervals are of magnitude f_1, the second and last but one are f_2, and so on; if we take five intervals in all, the centre one will be denoted by f_3. Then, as in § 13 (4), we write

$$F'_0(t)=\sum_{1,2} f_m(e^{-F_m t}+e^{-(1-F_m)t})+f_3 e^{-\frac{1}{2}t}. \quad . \quad . \quad . \quad . \quad (2)$$

Thus we have five unknowns, f_1, f_2, f_3, F_1 and F_2. By differentiating (2) successively n times we have

$$F'_n(t)=\sum_{1,2} f_m\{F_m^n e^{-F_m t}+(1-F_m)^n e^{(1-F_m)t}\}+(f_3/2^n)e^{\frac{1}{2}t} \quad . \quad . \quad (3)$$

as an index-sum for

$$F_n(t)=\int_9^1 e^{-tx}x^n\,dx. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Clearly $F(0)=1/(n+1)$; we determine the five unknown constants in (2, 3) by making $F'_n(0)$ approximately equal to $F_n(0)$ for five values of n ranging from 0 to 10. Then by taking $e_m=f_m$ and $\epsilon_m=1/F_m$ or $\epsilon_m=1/(1-F_m)$ we obtain $E'_n(t)$.

15. An alternative *four*-term index-sum for $E_n(t)$, which we shall denote by $E''_2(t)$, is given by the following table (*cf.* § 3 (4)):—

m	1	2	3	4
$e_m=$	0·1521	0·3313	0·4028	0·1138
$\epsilon_m=$	1·0637	1·436	3·212	40·12

The first three terms in E''_n were found by making $E''_n(0)$ approximately equal to $E_n(0)$, that is, to $1/(n-1)$, for $n=11$ and lower values. The last index-coefficient e_4 was then determined to satisfy $E''_2(0)=E_2(0)=1$; the index-factor ϵ_4 was determined by considering the associated approximation

$$\Sigma e_m e^{-t\sqrt{\epsilon_m}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

for

$$\int_1^\infty e^{-t\sqrt{x}}x^{-2}\,dx. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

By differentiating (1) with respect to t, and then taking $t=0$, we obtain the approximation $\Sigma e_m\sqrt{\epsilon_m}$ for

$$\int_1^\infty x^{-3/2}\,dx,$$

whose value is 2. Knowing every e_m, and every ϵ_m except ϵ_4, this gives ϵ_4.

For the values $n=0$ to $n=7$ the error $\Delta E''_n(0)$ for the index-sum $E''_n(t)$, for $t=0$, is as follows:—$10^{-4}\times(0, 6, 0, 0, 2, 3, 1, 1)$, so that at least for $t=0$ this index-sum is better than the five-term formula given in § 4.

For numerical comparison of our present index-sum for non-zero values of t, we consider $e^t\mathrm{E}_1(t)$, which is given by § 1 (5) for large values of t; for $t=1$ its value is given by Knopp (1929) as 0·596347. The following table shows that $e^t\mathrm{E}_1''(t)$ agrees with $e^t\mathrm{E}_1(t)$ within $\frac{1}{2}$ per cent. when $t=1$ (in this case § 1 (4) is not applicable) and within 0·0002 for values of t as large as 10 or 15.

$t=$	1	10	15
Integral (3)......	0·5963	0·0914	0·0628
Index-sum (2) ...	0·5982	0·0916	0·0629

Of course, for very large values of t the index-sum is not valid, and § 1 (5) is then both accurate and easy to compute :

16. Many other examples of index-sums might be given, including some for elliptic functions : the London Ph.D. thesis from which most of this paper is taken contains further details of such index-sums, and some general discussion of this method of numerical approximation. There is need, however, for further investigation of the method, from the standpoint of pure mathematics.

17. As a final example, we give an index-sum for an integral considered by E. V. Appleton, in the discussion (in his Bakerian lecture) of the electrical conductivity of the ionosphere ; namely

$$\mathrm{G}(t)\equiv\int_0^\infty \frac{e^{-tx}\sqrt{x}}{1+x^2}dx. \quad \cdots \quad (1)$$

From the index-sum approximation to $\mathrm{E}_n(t)$, using the integral analogy indicated in § 6, we obtain the index-sum

$$\mathrm{G}'(t)=\Sigma e_m\epsilon_m^2(1-\epsilon_m)^{1/2}e^{(1-\epsilon_m)t}\equiv\Sigma g_m e^{-\gamma_m t}. \quad \cdots \quad (2)$$

Adopting a four-term formula and the values of e_m and ϵ_m given in § 15, we obtain the following constants in (2) :—

$m=$	1	2	3	4
$g_m=$	0·04326	0·37911	1·05181	0·7485
$\gamma_m=$	0·06371	0·4363	2·2425	39·16

It is not difficult to establish the following series for $\mathrm{G}(t)$, for small and for large values of t respectively :

$$\mathrm{G}(t)=\frac{\pi}{\sqrt{2}}(\cos t+\sin t)-2(\pi t)^{1/2}\left\{1-\frac{(2t)^2}{3.5}+\frac{(2t)^4}{3.5.7.9}\cdots\right\}. \quad . \quad (3)$$

$$\mathrm{G}(t)=\frac{\sqrt{\pi}}{2t^{3/2}}\left\{1-\frac{3.5}{(2t)^2}+\frac{3.5.7.9}{(2t)^4}-\cdots\right\}. \quad \cdots \quad (4)$$

From (3) we have calculated $\mathrm{G}(t)$ for four values of t, namely 0, 10^{-2}, 10^{-1}, and 1, with the results which in the following table are compared with the corresponding values of $G'(t)$:—

$t=$	0	0·01	0·1	1
$G(t)=$	2·221	1·99	1·28	0·41
$G'(t)=$	2·223	1·96	1·26	0·40

The values $G'(t)$ are in this case amply accurate enough for practical purposes.

17. The practical convenience of index-sum approximations is particularly well illustrated by reference to the theory of the multiple scattering of light in a plane-stratified atmosphere (Hammad and Chapman, 1939). The consideration of the *emission* of secondary scattered light introduces functions L_n, which involve integrals of the type (*loc. cit.* p. 105)

$$Ek_n(T, b) \equiv \int_0^T e^{bt} E_n(t)\, dt. \qquad (1)$$

By the use of the formulæ given in § 1, these integrals can be expressed in terms of the integral E_1 and of logarithmic functions ; however, the formulæ become increasingly complicated as n increases. But by substituting in (1) an index-sum approximation for $E_n(t)$, the integration can easily be effected, the result being another series of exponential terms. Since either b is negative, or b is positive and the integral Ek_n is multiplied by e^{-bT}, the functions L_n can themselves be expressed in the form of index-sums, whose index-factors exceed those of the corresponding terms in E_n'.

The *reception* of secondary scattered light introduces further functions Lu_n and Ld_n (*loc. cit.*, p. 106), which are of the form

$$\int_0^T e^{-ct} L_n(t)\, dt,$$

where c is negative. By substituting the index-sum approximations to L_n, these further functions can likewise be expressed as index-sums. The consideration of tertiary and higher orders of scattered light leads to still more complicated integrals that can be similarly treated.

In numerical calculations it is possible at each stage to drop any terms in the index-sums, whose index-factor has, by successive addition, become so large that the term is negligible.

References.

Appleton, E. V., Proc. Roy. Soc. A, clxii. p. 451 (1937).
Dale, J. B., 'Five-figure Tables of Mathematical Functions.' London, Arnold (1903).
Glaisher, J. W. L., Phil. Trans. Roy. Soc. clx. (1878).

Gold, E., Proc. Roy. Soc. A, lxxxii. p. 43 (1909).
Hammad, A., and Chapman, S., Phil. Mag. xxviii. p. 99 (1939).
Hopf, E., 'Mathematical Problems of Radiative Equilibrium,' Cambridge Tracts, No. 31 (1934).
Jahnke, E., and Emde, F., 'Funktionentafeln.'
Knopp, K., 'Theory and Application of Infinite Series' (English translation, p. 555) (1929).
Nielsen, N., 'Theorie des Integrallogarithmus,' Leipzig (1906).
Roberts, O. F. T., Proc. Roy. Soc., Edin. l. p. 225 (1930).
Unsöld, A., 'Physik der Sternatmosphären,' App. B., p. 467 (1938).

XV. *The Forces on an Electron, according to Dirac's Wave Equation.*

By A. Lees, M.Sc., Ph.D., University College, Swansea *.

[Received December 4, 1941.]

Summary.

It is proved that the forces on an electron, according to Dirac's wave equation, consist of a purely internal system of stresses together with those forces which would be exerted upon it by the external electromagnetic field according to classical principles.

1. Various fundamental deductions have been made from the linear wave equation for the electron introduced by Dirac †, for example the existence of the electron spin and (in a sense) of the magnetic moment of the electron. There is, however, one equally important point which does not seem to have been investigated, namely, exactly what forces act upon the electron, for the answer to this is not obvious from the form of the equation. We therefore propose now to determine the system of forces which acts on the electron.

2. We first establish our notation regarding Dirac's equation, and some results which we shall need. As all of the formulæ in this section are either well known or easily obtained, we shall state them without proof.

We write Dirac's equation in the form

$$E^{\mu}\psi_{\mu}=\left(\frac{m}{\hbar}+\frac{ie}{\hbar}E^{\mu}\phi_{\mu}\right)\psi. \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.1)$$

Here, as henceforth, we adopt the tensor convention of summing repeated suffixes over the values, 1, 2, 3, 4. ψ is the wave function, a function of rectangular space co-ordinates x^1, x^2, x^3, and time x^4. ψ_μ denotes $\frac{\partial\psi}{\partial x^\mu}$, the covariant derivative of ψ, ϕ_μ is the potential vector for the external electromagnetic field ‡, m is the mass of the electron, $-e$ is its charge, and $\hbar$ denotes $\frac{h}{2\pi}$, h being the quantum of action. We take the unit of time so that the velocity of light is unity; thus the metric is given by

---

* Communicated by Professor H. T. H. Piaggio, M.A., D.Sc.

† Proc. Roy. Soc. A, vol. cxvii. p. 610 (1928).

‡ $\phi_\mu \equiv (-A_1, -A_2, -A_3, V)$, where A_1, A_2, A_3 are vector potential components and V is the scalar potential.

$$ds^2=g_{\mu\nu}\,dx^\mu\,dx^\nu=-\sum_{l=1}^{3}(dx^l)^2+(dx^4)^2.$$

The E^μ's are fixed matrices which satisfy the equations

$$E^\mu E^\nu+E^\nu E^\mu=-2g^{\mu\nu}, \qquad \qquad (2.2)$$

and E^1, E^2, E^3, and iE^4 are Hermitian. Except in the case of the E^μ's we shall always use high and low suffixes to denote respectively contravariant and covariant vectors or tensors.

It may be verified that (2.1) agrees with the standard form of the linear wave equation *.

Let a Lorentz transformation be applied, the variables x^μ changing to new variables $(x^\mu)'$. Let us denote by $(E^\mu)'$ the matrices obtained by supposing the E^μ's to transform as a contravariant vector, in formal agreement with their high suffix μ. Then there exists a matrix T such that

$$(E^\mu)'=TE^\mu T^{-1}, \qquad \qquad (2.3)$$

and we restrict T by the condition

$$\text{mod.}\,|T|=1, \qquad \qquad (2.4)$$

$|T|$ being the determinant of T. If ψ now be supposed to transform to ψ', given by

$$\psi'=T^{-1}\psi,$$

then (2.1) and the normalization of ψ are invariant for Lorentz transformations.

Using bars to denote conjugate complexes of transposed matrices,

$$\overline{E^l}=E^l,\ l=1,2,3,\ \overline{E^4}=-E^4,\ \overline{E^\mu}E^4=-E^4E^\mu. \qquad \qquad (2.5)$$

This last may also be written

$$\overline{E^4E^\mu}=E^4E^\mu,$$

so that E^4E^μ is Hermitian. Hence $\bar{\psi}E^4E^\mu\psi$ is a real quantity.

For T as in (2.3) and (2.4),

$$\overline{T^{-1}}E^4=E^4T.$$

As a result of this, it may be proved that such quantities as $\bar{\psi}E^4E^\nu \ldots E^\kappa\psi$, $\bar{\psi}_\mu E^4E^\nu \ldots E^\kappa\psi$, $\bar{\psi}E^4E^\nu \ldots E^\kappa\psi_\mu$, are all tensors. In particular, $\bar{\psi}E^4E^\mu\psi$ is a vector. We write

$$-\bar{\psi}E^4E^\mu\psi=\rho^\mu. \qquad \qquad (2.6)$$

---

* See first footnote.

Then we have seen above that ρ^μ is real, and in fact ρ^μ is the flux vector for the electron and satisfies the equation

$$\rho^\mu_\mu = 0. \qquad (2.7)$$

ϕ_μ satisfies the equation

$$\phi^\mu_\mu = 0. \qquad (2.8)$$

and the field tensor $F_{\mu\nu}$ is defined by

$$F_{\mu\nu} = \phi_{\mu\nu} - \phi_{\nu\mu}. \qquad (2.9)$$

The energy H of the electron is equivalent to the operator $ih\dfrac{\partial}{\partial x^4}$ when applied to ψ.

3. To investigate the forces on the electron, we shall set up a stress-momentum-energy tensor $T^{\mu\nu}$ for it. $T^{\mu\nu}$ is to correspond to the electron only, and to involve nothing attributable to the external field.

Thus T^{44} must represent the spatial density of the kinetic energy $H + e\phi_4$, the whole energy H less the energy $-e\phi_4$ due to the field. As $H + e\phi_4$ is equivalent to the operator $ih\dfrac{\partial}{\partial x^4} + e\phi_4$ applied to ψ, we can as a first attempt put T^{44} equal to $\bar\psi(i\hbar\psi_4 + e\phi_4\psi)$, or, as the raising of a suffix 4 makes no difference, to $\bar\psi(i\hbar\psi^4 + e\phi^4\psi)$. We are in agreement with this if we put $T^{\mu\nu}$ equal to $-\bar\psi E^4 E^\mu(i\hbar\psi^\nu + e\phi^\nu\psi)$, which is a tensor. This tensor, however, is neither real nor symmetrical. We make it symmetrical by interchanging μ and ν and taking half of the sum of the two expressions. and make it real by taking its real part, denoted by R. This gives, by (2.6),

$$T^{\mu\nu} = \tfrac{1}{2}R[-i\hbar(\bar\psi E^4 E^\mu \psi^\nu + \bar\psi E^4 E^\nu \psi^\mu) + e(\rho^\mu\phi^\nu + \rho^\nu\phi^\mu)]. \qquad (3.1)$$

4. We now take the contracted covariant derivative of (3.1), giving

$$(T^{\mu\nu})_\nu = \tfrac{1}{2}R[-i\hbar(\bar\psi_\nu E^4 E^\mu \psi^\nu + \bar\psi E^4 E^\mu \psi^\nu_\nu + \bar\psi_\nu E^4 E^\nu \psi^\mu + \bar\psi E^4 E^\nu \psi^\mu_\nu) \\ + e(\rho^\mu{}_\nu\phi^\nu + \rho^\mu\phi^\nu_\nu + \rho^\nu_\nu\phi^\mu + \rho^\nu\phi^\mu{}_\nu)]. \qquad (4.1)$$

As E^4E^μ is Hermitian, it is easily seen that $\bar\psi_\nu E^4 E^\mu \psi^\nu$ is real, so the first term on the right of (4.1) vanishes. The last term but two vanishes by (2.8), and the last term but one vanishes by (2.7).

For the second term on the right of (4.1), we use the relation, derivable from (2.2),

$$\psi^\nu_\nu = -E^\mu \frac{\partial}{\partial x^\mu}(E^\nu\psi_\nu).$$

By (2.1), this gives

$$\psi^\nu_\nu = -E^\mu \frac{\partial}{\partial x^\mu}\left(\frac{m}{\hbar} + \frac{ie}{\hbar}E^\nu\phi_\nu\right)\psi.$$

or, again using (2.1),

$$\psi_\nu^\nu = -\frac{m}{\hbar}\left(\frac{m}{\hbar} + \frac{ie}{\hbar}\mathrm{E}^\mu\phi_\mu\right)\psi - \frac{ie}{\hbar}\mathrm{E}^\mu\mathrm{E}^\nu\psi\phi_{\nu\mu} - \frac{ie}{\hbar}\mathrm{E}^\mu\mathrm{E}^\nu\psi_\mu\phi_\nu.$$

In the last term of this expression we again use (2.2), giving

$$\psi_\nu^\nu = -\frac{m^2}{\hbar^2}\psi - \frac{ime}{\hbar^2}\mathrm{E}^\mu\phi_\mu\psi - \frac{ie}{\hbar}\mathrm{E}^\mu\mathrm{E}^\nu\psi\phi_{\nu\mu} + \frac{ie}{\hbar}\mathrm{E}^\nu\mathrm{E}^\mu\psi_\mu\phi_\nu + \frac{2ie}{\hbar}\psi_\mu\phi^\mu.$$

To the last term but one of this expression we apply (2.1) again, obtaining two terms, one of which cancels the second term above, and we get

$$\psi_\nu^\nu = -\frac{m^2}{\hbar^2}\psi - \frac{ie}{\hbar}\mathrm{E}^\mu\mathrm{E}^\nu\psi\phi_{\nu\mu} + \frac{2ie}{\hbar}\psi_\mu\phi^\mu - \frac{e^2}{\hbar^2}\mathrm{E}^\nu\mathrm{E}^\mu\psi\phi_\mu\phi_\nu. \quad . \quad . \quad (4.2)$$

In the second term of (4.2) we interchange μ and ν and take half of the sum of the two quantities. This gives

$$-\frac{ie}{\hbar}\mathrm{E}^\mu\mathrm{E}^\nu\psi\phi_{\nu\mu} = -\frac{ie}{2\hbar}(\mathrm{E}^\mu\mathrm{E}^\nu\psi\phi_{\nu\mu} + \mathrm{E}^\nu\mathrm{E}^\mu\psi\phi_{\mu\nu})$$

$$= -\frac{ie}{2\hbar}[(\mathrm{E}^\mu\mathrm{E}^\nu + \mathrm{E}^\nu\mathrm{E}^\mu)\psi\phi_{\nu\mu} + \mathrm{E}^\nu\mathrm{E}^\mu\psi(\phi_{\mu\nu} - \phi_{\nu\mu})].$$

Using (2.2), (2.8), and (2.9), this reduces to

$$-\frac{ie}{\hbar}\mathrm{E}^\mu\mathrm{E}^\nu\psi\phi_{\nu\mu} = -\frac{ie}{2\hbar}\mathrm{E}^\nu\mathrm{E}^\mu\psi\mathrm{F}_{\mu\nu}. \quad . \quad . \quad . \quad . \quad . \quad (4.3)$$

In the last term of (4.2) we again interchange μ and ν and take half the sum of the two quantities. This gives

$$-\frac{e^2}{\hbar^2}\mathrm{E}^\nu\mathrm{E}^\mu\psi\phi_\mu\phi_\nu = -\frac{e^2}{2\hbar^2}(\mathrm{E}^\nu\mathrm{E}^\mu + \mathrm{E}^\mu\mathrm{E}^\nu)\psi\phi_\mu\phi_\nu,$$

which, by (2.2), reduces to

$$\frac{e^2}{\hbar^2}g^{\mu\nu}\psi\phi_\mu\phi_\nu = \frac{e^2}{\hbar^2}\psi\phi_\mu\phi^\mu. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.4)$$

Substituting from (4.3) and (4.4) in (4.2), we finally have that

$$\psi_\nu^\nu = -\frac{m^2}{\hbar^2}\psi - \frac{ie}{2\hbar}\mathrm{E}^\nu\mathrm{E}^\mu\psi\mathrm{F}_{\mu\nu} + \frac{2ie}{\hbar}\psi_\mu\phi^\mu + \frac{e^2}{\hbar^2}\psi\phi_\mu\phi^\mu. \quad . \quad . \quad . (4.5)$$

For the third term of (4.1), we use (2.5), giving

$$\bar{\psi}_\nu\mathrm{E}^4\mathrm{E}^\nu = -\bar{\psi}_\nu\overline{\mathrm{E}}^\nu\mathrm{E}^4 = -(\overline{\mathrm{E}^\nu\psi_\nu})\mathrm{E}^4,$$

or, using (2.1),

$$\bar{\psi}_\nu\mathrm{E}^4\mathrm{E}^\nu = -\bar{\psi}\left(\frac{m}{\hbar} - \frac{ie}{\hbar}\overline{\mathrm{E}}^\nu\phi_\nu\right)\mathrm{E}^4.$$

Again using (2.5), this becomes

$$\bar{\psi}_\nu\mathrm{E}^4\mathrm{E}^\nu = -\bar{\psi}\mathrm{E}^4\left(\frac{m}{\hbar} + \frac{ie}{\hbar}\mathrm{E}^\nu\phi_\nu\right). \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.6)$$

For the fourth term of (4.1) we use (2.1), giving

$$\mathrm{E}^{\nu}\psi_{\nu}^{\mu}=(\mathrm{E}^{\nu}\psi_{\nu})^{\mu}=\left\{\left(\frac{m}{\hbar}+\frac{ie}{\hbar}\,\mathrm{E}^{\nu}\phi_{\nu}\right)\psi\right\}^{\mu},$$

or

$$\mathrm{E}^{\nu}\psi_{\nu}^{\mu}=\left(\frac{m}{\hbar}+\frac{ie}{\hbar}\,\mathrm{E}^{\nu}\phi_{\nu}\right)\psi^{\mu}+\frac{ie}{\hbar}\,\mathrm{E}^{\nu}\psi\phi_{\nu}{}^{\mu}. \qquad (4.7)$$

We now substitute from (4.5), (4.6), and (4.7) in (4.1), omitting the vanishing terms and four terms which cancel in pairs, and using (2.6). This gives

$$(\mathrm{T}^{\mu\nu})_{\nu}=\tfrac{1}{2}\mathrm{R}\left[-\frac{im^{2}}{\hbar}\,\rho^{\mu}+e\left(-\tfrac{1}{2}\bar{\psi}\mathrm{E}^{4}\mathrm{E}^{\mu}\mathrm{E}^{\nu}\mathrm{E}^{\sigma}\psi\mathrm{F}_{\sigma\nu}+2\bar{\psi}\mathrm{E}^{4}\mathrm{E}^{\mu}\psi_{\nu}\phi^{\nu}\right.\right.$$
$$\left.\left.+\frac{ie}{\hbar}\,\rho^{\mu}\phi_{\nu}\phi^{\nu}-\rho^{\nu}\phi_{\nu}{}^{\mu}+\rho^{\mu}{}_{,\nu}\phi^{\nu}+\rho^{\nu}\phi^{\mu}{}_{,\nu}\right)\right]. \qquad (4.8)$$

The first and fourth terms on the right of (4.8) vanish when their real parts are taken. For the third term, as $\mathrm{E}^{4}\mathrm{E}^{\mu}$ is Hermitian it is easily seen that

$$\mathrm{R}(2\bar{\psi}\mathrm{E}^{4}\mathrm{E}^{\mu}\psi_{\nu})=(\bar{\psi}\mathrm{E}^{4}\mathrm{E}^{\mu}\psi)_{\nu}=-\rho^{\mu}{}_{,\nu},$$

and this causes the third and sixth terms of (4.8) to cancel. Also the fifth and seventh terms combine by (2.9). With these simplifications, (4.8) becomes

$$(\mathrm{T}^{\mu\nu})_{\nu}=\frac{e}{2}\,\mathrm{R}(-\tfrac{1}{2}\bar{\psi}\mathrm{E}^{4}\mathrm{E}^{\mu}\mathrm{E}^{\nu}\mathrm{E}^{\sigma}\psi\mathrm{F}_{\sigma\nu}+\rho^{\nu}\mathrm{F}^{\mu}{}_{\nu}). \qquad (4.9)$$

To simplify the first term of (4.9), we observe that it is a vector, so it is sufficient to evaluate any one component, so long as we express the result as a vector component. Taking μ equal to 4, the term becomes

$$-\frac{e}{4}\,\mathrm{R}(\bar{\psi}\mathrm{E}^{4}\mathrm{E}^{4}\mathrm{E}^{\nu}\mathrm{E}^{\sigma}\psi\mathrm{F}_{\sigma\nu})=\frac{e}{4}\,\mathrm{R}(\bar{\psi}\mathrm{E}^{\nu}\mathrm{E}^{\sigma}\psi\mathrm{F}_{\sigma\nu}).$$

It is easily seen that, if $\nu\neq\sigma$, $\bar{\psi}\mathrm{E}^{\nu}\mathrm{E}^{\sigma}\psi$ will be imaginary unless one suffix has the value 4, and then it will be real. Hence the quantity is equal to

$$\frac{e}{4}\sum_{\nu=1}^{3}(\bar{\psi}\mathrm{E}^{4}\mathrm{E}^{\nu}\psi\mathrm{F}_{\nu4}+\bar{\psi}\mathrm{E}^{\nu}\mathrm{E}^{4}\psi\mathrm{F}_{4\nu}),$$

and by (2.2) and (2.6) this is the same as

$$-\frac{e}{2}\sum_{\nu=1}^{3}\rho^{\nu}\mathrm{F}_{\nu4}=-\frac{e}{2}\,\rho^{\nu}\mathrm{F}_{\nu}{}^{4}.$$

This result is in agreement with the vector equation

$$-\frac{e}{4}\,\mathrm{R}(\bar{\psi}\mathrm{E}^{4}\mathrm{E}^{\mu}\mathrm{E}^{\nu}\mathrm{E}^{\sigma}\psi\mathrm{F}_{\sigma\nu})=-\frac{e}{2}\,\rho^{\nu}\mathrm{F}_{\nu}{}^{\mu}, \qquad (4.10)$$

which is therefore true.

Substituting from (4.10) in (4.9) and simplifying, we have finally the equation

$$(T^{\mu\nu})_\nu = e\rho^\nu F^\mu_\nu . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.11)$$

5. Integrating (4.11) throughout any volume we have, in the manner of Green's theorem, the relation

$$\int \frac{\partial T^{\mu 4}}{\partial x^4} d_\tau = e\int \rho^\nu F^\mu_\nu d_\tau - \int T^{\mu n}\, dS, \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.1)$$

in which d_τ is an element of volume, dS is an element of the bounding surface, and n denotes the direction of the outward normal at dS.

The rate of change of the integral $\int T^{\mu 4} d_\tau$, when the boundary moves with the electron, is given by

$$\frac{D}{Dx^4}\int T^{\mu 4}\, d_\tau = \int \frac{\partial T^{\mu 4}}{\partial x^4} d_\tau + \int T^{\mu 4} v^n dS, \quad . \quad . \quad . \quad . \quad (5.2)$$

in which v^n is the normal outward velocity at dS. This velocity, we must note, is not to be taken as that of the charge element at dS, which is $\frac{\rho^n}{\rho^4}$, but as that of the element of mass, or of kinetic energy, which is the same thing. That is,

$$v^n = \frac{T^{4n}}{T^{44}} . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.3)$$

Hence, by (5.1), (5.2), and (5.3),

$$\frac{D}{Dx^4}\int T^{\mu 4}\, d_\tau = e\int \rho^\nu F^\mu_\nu d_\tau + \int \left(\frac{T^{\mu 4}T^{4n}}{T^{44}} - T^{\mu n}\right) dS. \quad . \quad . \quad . \quad (5.4)$$

For μ equal to 1, the left side of (5.4) gives the rate of increase of x^1 momentum, so the right side gives the total force in the direction of x^1 on this part of the electron. The surface integral shows that there is a tensile stress in the x^1 direction given by $\frac{T^{14}T^{4n}}{T^{44}} - T^{1n}$. This depends only on $T^{\mu\nu}$, so the system of stresses (we get similar results for x^2 and x^3) is purely internal, has no connexion with the external field, and has no effect on the electron as a whole because it is in equilibrium by itself.

The volume integral shows that there is a force per unit charge on the electron, in the direction of x^1, of amount $\frac{e\rho^\nu F^1_\nu}{(-e\rho^4)}$, or $F_{1\nu}\frac{\rho^\nu}{\rho^4}$. As $\frac{\rho^\nu}{\rho^4}$, for $\nu=1, 2, 3$, is the charge velocity in the direction of x^ν, this is just the classical expression for the force per unit charge.

Putting μ equal to 4 in (5.4), the left side becomes equal to the rate of increase of kinetic energy of this part of the electron. The right side therefore gives the rate at which work is done on this part of the electron by the acting forces. The surface integral now vanishes, showing that the stresses do no work*. The volume integral shows that there is work done on the electron of an amount per unit charge given by $\frac{e\rho^\nu F^4{}_\nu}{(-e\rho^4)}$, or $F_{\nu 4}\frac{\rho^\nu}{\rho^4}$. This again is just the classical expression for the rate of doing work per unit charge.

We conclude that the forces on the electron consist of a purely internal system of stresses, together with just those forces which would be exerted upon it by the external field according to classical principles.

I wish to express my thanks to Professor H. T. H. Piaggio for many helpful suggestions in connexion with this paper.

---

* This is of course an automatic consequence of our regarding the components $T^{\mu 4}$ as consisting wholly of the density of material momentum and kinetic energy. This view is really a mere convention.

XVI. *On Yule's Method of Investigating Periodicities of Disturbed Series. The Motion of a Pendulum in a Turbulent Fluid.*

By T. E. W. Schumann, Ph.D.*

[Received December 11, 1941.]

1. *Introduction.*

The harmonic analysis of a time series by Schuster's method has in general yielded disappointing results. For example, in the analysis of records of atmospheric pressure, numerous authors have detected many " significant " periodicities, but most of the periodicities thus discovered are not persistent and hence have proved to be of little, if any, practical value.

In an important paper Yule (1927) approached the problem from a new angle by introducing the idea of " disturbed " harmonic oscillations, which he visualises in the following words :—" If we observe at short equal intervals of time the departures of a simple harmonic pendulum from its position of rest, errors of observation will cause *superposed* fluctuations. . . . . The recording apparatus is left to itself, and unfortunately boys get into the room and start pelting the pendulum with peas, sometimes from one side, and sometimes from the other. The motion is now affected, not by *superposed fluctuations*, but by true *disturbances*, and the effect on the graph (showing the motion of the pendulum) will be of an entirely different kind. The graph will remain surprisingly smooth, but amplitude and phase will vary continually."

Yule then proceeded to investigate the motion of a pendulum subject to such random disturbances and applied his results to the analysis of Wolfer's sunspot numbers. His work was later extended by Sir Gilbert Walker (1931) and others, but thus far only the case where the disturbances are entirely random has been dealt with, and if any fundamental advance is to be made on Yule's original research, it becomes imperative to deal with the more general case of disturbances which are interrelated. The object of this paper then is to discuss the general case of disturbed harmonic motion, where the time series constituting the disturbances is not entirely haphazard, but shows finite coefficients of auto-correlation for finite time-lags.

---

* Communicated by the Author.

With this object in view, it seems advisable to deviate somewhat from Yule's rather formal treatment of the problem and to lay more stress on its dynamical aspects. Not only will the solution of the problem thereby be facilitated, but we shall also more readily be led to the correct interpretation of the statistical results.

2. *Statement of the Problem.*

Suppose we have a simple pendulum, subject to a certain amount of damping, which at *equal* intervals of time is bombarded in its line of motion by pellets all of equal mass, the velocities of the pellets being distributed according to the normal (Gaussian) or to any other symmetrical distribution law. Also suppose that the velocities of the pellets are not entirely random with respect to the time, but have given auto-correlations for given time-lags. Our problem is to investigate the motion of the pendulum in the statistical sense, *i. e.*, to express the parameters defining the average motion of the pendulum in terms of the parameters by which the motion of the pellets are described.

For this purpose we make use of the following notation :—

t=time.
ξ=time-lag.
h=interval of time between one impact and the next.
M=mass of the simple pendulum.
mh=mass of a single pellet.

$$\mathrm{D}=\frac{\mathrm{M}-mh}{\mathrm{M}+mh}.$$

$$\mathrm{E}=\frac{mh}{\mathrm{M}+mh}.$$

$y_0 \ldots . y_n$=displacement of the pendulum from its zero position at times $t=0, h, 2h \ldots . nh$.
$u_0{}' \ldots . u_n{}'$=the corresponding velocities of M just *before* impact.
$u_0 \ldots . u_n$= ,, ,, ,, *after* ,,
$2\pi/\alpha$=the (damped) period of the pendulum.
l=its damping factor.
$\lambda=l+m/\mathrm{M}$.
$\mathrm{Q}=e^{-lh}$.
$v_1 \ldots . v_n$=the velocities of the first to the nth pellet.
$\mathrm{R}_1 \ldots . \mathrm{R}_s$=coefficients of auto-correlation of the v's after time-lags of $h \ldots . sh$.
$r_1 \ldots . r_s$=the corresponding auto-correlations of the y's.

$M^2\eta^2=\frac{m^2}{n}\Sigma v_r{}^2=$the sum of the squares of the momenta of the pellets striking the pendulum in unit time.

Eventually we shall have occasion to consider the limiting case, where the interval of time h becomes indefinitely small, in which case the kinetic energy of the pellets striking the pendulum would become infinitely large if the mass of a single pellet remained finite. This is the reason why we have assumed the mass of the pellet proportional to h, for then the energy of the pellets striking the pendulum per unit time remains constant even though h is varied at will.

Specifically our problem is to derive relationships between the motion of the pendulum and the original disturbances producing that motion.

3. *Derivation of the Fundamental Equations.*

In the first place we can obtain an expression between u_r' and u_r, the velocities of the pendulum just before and just after the rth impact, by applying the laws of the conservation of momentum and of energy and supposing the impacts to be perfectly elastic. The relation is

$$u_r=\frac{M-mh}{M+mh}u_r'+\frac{2mh}{M+mh}v_r,$$
$$=Du_r'+2Ev_r. \qquad (1)$$

Between impacts the equation of motion of the pendulum is

$$\frac{d^2y}{dt^2}-2l\frac{dy}{dt}+(d^2+l^2)y=0, \qquad (2)$$

and the solution of this equation may be written

$$y=Ae^{-lt}\sin(\alpha t+\omega)$$

and $$u=-lAe^{-lt}\sin(\alpha t+\omega)+\alpha Ae^{-lt}\cos(\alpha t+\omega).$$

In particular we may write

$$y_0=A_1\sin\omega_1. \qquad (3)$$
$$u_0=-lA_1\sin\omega_1+\alpha A_1\cos\omega_1. \qquad (4)$$
$$y_1=A_1Q\sin(\alpha h+\omega_1). \qquad (5)$$
$$u_1'=-lA_1Q\sin(\alpha h+\omega_1)+\alpha A_1Q\cos(\alpha h+\omega_1). \qquad (6)$$

Now, due to the impact at the time $t=h$, both the amplitude and the phase are changed to A_2 and ω_2 respectively (say). Accordingly, since y_1 is not instantaneously changed by the impact, we may again write

$$y_1=A_2\sin\omega_2. \qquad (7)$$
$$u_1=-lA_2\sin\omega_2+\alpha A\cos\omega_2. \qquad (8)$$
$$y_2=A_2Q\sin(\alpha h+\omega_2).$$

But according to (1)

$$u_1 = \mathrm{D}u_1' + 2\mathrm{E}v_1$$

$$= -\mathrm{D}l\mathrm{A}_1\mathrm{Q}\sin(\alpha h+\omega_1) + \mathrm{D}\alpha\mathrm{A}_1\mathrm{Q}\cos(\alpha h+\omega_1) + 2\mathrm{E}v_1. \quad . \quad . \quad (9)$$

Between equations (3) to (9) it is possible to eliminate the six constants A_1, A_2, ω_1, ω_2, u_0 and u_1, the result being

$$y_2 = kqy_1 - q^2y_0 + \epsilon_2,$$

or, in general,

$$y_r = kqy_{r-1} - q^2y_{r-2} + \epsilon_r, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where we have introduced the constants k, q, and ϵ_r, their values being

$$kq = \mathrm{Q}(\mathrm{D}+1)\cos\alpha h + \frac{\mathrm{Q}l}{\alpha}(1-\mathrm{D})\sin\alpha h. \quad . \quad . \quad . \quad . \quad (11)$$

$$q^2 = \mathrm{DQ}^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

$$\epsilon_r = 2\mathrm{QE}v_{r-1}\frac{\sin\alpha h}{\alpha}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

We may also write down the corresponding equation for u_r, derived in exactly the same way :

$$u_r = kqu_{r-1} - q^2u_{r-1} + \delta_r, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

where

$$\delta_r = 2\mathrm{E}\left\{v_r - \mathrm{Q}v_{r-1}\left(\frac{l}{\alpha}\sin\alpha h + \cos\alpha h\right)\right\}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

As we shall, in all our subsequent work, consider n to be very large, the series being practically infinite in extent, we may without loss of generality put y_0 equal to zero, in which case it follows from (10) that

$$y_1 = \epsilon_1,$$

$$y_2 = kq\epsilon_1 + \epsilon_2,$$

$$y_3 = (k^2q^2 - q^2)\epsilon_1 + kq\epsilon_2 + \epsilon_r; \text{ etc.}$$

In general these equations may be written

$$y_r = \phi_0\epsilon_r + \phi_1\epsilon_{r-1} + \dots\dots \phi_{r-1}\epsilon_1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where we have introduced the function ϕ_r, whose value is given by

$$\phi_0 = 1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

$$\phi_1 = kq, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

and

$$\phi_{r+1} = kq\phi_r - q^2\phi_{r-1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

The last is a difference equation which we shall encounter again in the course of this work. Accordingly, we briefly show how its solution may be obtained ;

Subtract $p\phi_r$ from both sides of (19)

$$\phi_{r+1}-p\phi_r=(kq-p)\left(\phi_r-\frac{q^2}{kq-p}\phi_{r-1}\right). \quad . \quad . \quad . \quad . \quad (20)$$

Now choose p such that $p^2-kqp+q^2=0$. . . . . (21)

Two cases arise according to whether k is less or greater than 2. If $k>2$ the motion of the pendulum corresponds to the highly damped aperiodic case. For the time being we leave this case out of consideration and confine our attention to the motion when $k<2$, in which case a solution of the quadratic (21) is

$$p=qe^{-i\theta}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

where $k=2\cos\theta$. . . . . (23)

It thus follows from (20) that

$$\begin{aligned}\phi_{r+1}-p\phi_r&=\frac{q^2}{p}(\phi_r-p\phi_{r-1})\\&=qe^{i\theta}(\phi_r-p\phi_{r-1}).\end{aligned}$$

Reducing the order of the right-hand side step by step, we finally find

$$\phi_{r+1}-e^{-i\theta}\phi_r=q^r e^{ir\theta}(\phi_1-e^{-i\theta}\phi_0). \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

Equating the imaginary terms of (24) we obtain the result

$$\phi_r\sin\theta=q^{r-1}\phi_1\sin r\theta-q^r\phi_0\sin(r+1)\theta. \quad . \quad . \quad . \quad . \quad (25)$$

This is the general solution of the difference equation (19), but in our particular case, where $\phi_0=1$ and $\phi_1=kq=2q\cos\theta$, the value of ϕ_r reduces to

$$\phi_r=q^r\sin(r+1)\theta/\sin\theta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

In equation (16), which may be re-written in the form

$$y_r=\sum_{s=0}^{r}\phi_s\epsilon_{r-s}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

the deviations $y_0 \ldots y_n$ are expressed in terms of the disturbances $\epsilon_1 \ldots \epsilon_n$ and of the functions ϕ_r, defined in (26). Equations (26) and (27) may therefore be regarded as fundamental to the further development of the theory.

4. *The Computation of the Auto-correlation Coefficients.*

We can now proceed to calculate the standard deviation of the y's, as well as the auto-correlation coefficients r_1, r_2, r_3, etc., and for this purpose we employ the symbols

$$\sigma_0{}^2=\frac{1}{n}\sum_{0}^{n}y_r{}^2,$$

$$\sigma_s{}^2=\frac{1}{n-s}\sum_{0}^{n-s}y_r y_{r+s}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

We also use the symbol σ, without subscript, to denote the standard deviation of the disturbances

$$\sigma^2 = \frac{1}{n}\sum_{0}^{n} \epsilon_r^{\,2}. \qquad (29)$$

By ordinary algebra it is now possible to determine the value of σ_s; *e. g.*, from (15) it is evident that

$$y_r y_{r+s} = (\phi_0\epsilon_r + \ldots \phi_r\epsilon_0)(\phi_0\epsilon_{r+s} + \ldots \phi_{r+s}\epsilon_0).$$

If the multiplication is carried out and the values of $y_r y_{r+s}$ for all values of r are added together, the different terms being properly grouped and due regard being paid to the fact that the ϕ's form a convergent series and that the series is practically infinite in extent, it is possible to demonstrate that

$$\sigma_0^{\,2} = \sigma^2(\chi_0 + 2\mathrm{R}_1\chi_1 + 2\mathrm{R}_2\chi_2 + \ldots),$$

and, in general,

$$\sigma_s^{\,2} = \sigma^2\{\chi_s + \mathrm{R}_1(\chi_{s-1} + \chi_{s+1}) + \ldots \mathrm{R}_s(\chi_0 + \chi_{2s}) + \mathrm{R}_{s+1}(\chi_1 + \chi_{2s+1}) + \ldots\}, \qquad (30)$$

where the R's are the autocorrelations of the ϵ's as previously stated, and where the functions χ_r are defined thus:

$$\left.\begin{aligned} \chi_0 &= \sum_{0}^{\infty}\phi_r^{\,2}, \\ \text{and} \qquad \chi_s &= \sum_{0}^{\infty}\phi_r\phi_{r+s} \end{aligned}\right\} \qquad (31)$$

Our next step is to determine the values of the χ's, for which purpose we make use of the following identity, an expression which may readily be derived by using the properties of complex quantities:

$$\sum_{x=0}^{\chi} a^x \sin(\omega x + \beta) = \frac{\sin\beta - a\sin(\beta-\omega) - a^{\chi+1}\sin(\omega\chi+\omega+\beta) + a^{\chi+2}\sin(\omega\chi+\beta)}{1 + a^2 - 2a\cos\omega}. \qquad (32)$$

If this equation is applied, the following results can be derived from (31) without much difficulty:

$$\chi_0 = \frac{1+q^2}{(1-q^2)\mathrm{L}^2}, \qquad (33)$$

and

$$\chi_1 = \frac{kq}{(1-q^2)\mathrm{L}^2}, \qquad (34)$$

where

$$\mathrm{L}^2 = (1+q^2)^2 - k^2q^2. \qquad (35)$$

Furthermore, we note that

$$\chi_s = kq\chi_{s-1} - q^2\chi_{s-2}. \qquad (36)$$

This is the same difference equation we encountered in (19), whose general solution is contained in (25). Thus, the final value of χ_s, after suitable transformation, is found to be

$$\chi_s=\frac{q^s \sin(s\theta+\gamma)}{(1-q^2)\mathrm{L}\sin\theta}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

where $\sin\gamma=\frac{1+q^2}{\mathrm{L}}\sin\theta$ and $\cos\gamma=\frac{1-q^2}{\mathrm{L}}\cos\theta$. . . . (38)

Equation (30) may now be written in the form

$$r_s=\frac{\sum\limits_{r=0}^{\infty}(\mathrm{R}_s\chi_{r+s}+\mathrm{R}_{r+s}\chi_r)+\sum\limits_{r=0}^{s-2}\mathrm{R}_{r+1}\chi_{s-r-1}}{2\sum\limits_{0}^{\infty}\mathrm{R}_r\chi_r-\chi_0}, \quad . \quad . \quad . \quad (39)$$

where $\sigma_0^2/\sigma^2=2\sum\limits_{0}^{\infty}\mathrm{R}_r\chi_r-\chi_0$. . . . . (40)

In (39) and (40) we have expressions for the standard deviation and the auto-correlations of the y's, and our problem has thus been formally solved. However, the result still remains in the form of a series, and if r_s is to be expressed in simpler analytic form, it is essential that R should also be given in analytic form, and thus we now proceed to discuss the particular case where R is an exponential function of the time. At the same time it should be pointed out that it is quite possible also to treat other cases, *e. g.*, where R itself is a damped simple harmonic function.

5. *Particular Case* : R *an Exponential Function of the Time-Lag.*

As shown in a previous publication by the present author (1941), we have good reason to devote special attention to the case where the auto-correlation coefficient decreases exponentially with the time. Accordingly we put

$$\mathrm{R}_1=e^{-\mu h}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

and

$$\mathrm{R}_s=e^{-\mu s h}=\mathrm{R}_1^s. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

From (39) and (40) it follows that

$$\sigma_s^2(1-q^2)\mathrm{L}\sin\theta/\sigma^2=q^s\sum_{0}^{\infty}\mathrm{R}_1{}^r q^r\sin(r\theta+s\theta+\gamma)+\mathrm{R}_1{}^s\sum_{0}^{\infty}\mathrm{R}_1{}^r q^r\sin(r\theta+\gamma)$$
$$-\mathrm{R}_1 q^{s-1}\sum_{0}^{s-2}\mathrm{R}_1{}^r q^{-r}\sin(r\theta+\theta-s\theta-\gamma). \quad (43)$$

The evaluation of this expression by the aid of (32) is fairly laborious, but it involves only straightforward algebra, and leads to the result

$$\frac{\mathrm{L}^2\mathrm{M}^2\mathrm{N}^2\sigma_s^2}{\sigma^2}=\frac{\mathrm{R}_1(1+q^2)(2-k^2)+kq(1+\mathrm{R}_1{}^2)}{(4-k^2)^{\frac{1}{2}}}(1-\mathrm{R}_1{}^2)q^{s+1}\sin s\theta$$

$$+\frac{q(1+R_1^2)(1+q^2)-R_1k(1+q^4)}{(1-q^2)}(1-R_1^2)q^{s+1}\cos s\theta$$

$$+N^2R_1^{s+2}, \qquad (44)$$

where
$$M^2=1+R_1^2q^2-R_1kq, \qquad (45)$$

$$N^2=R_1^2+q^2-R_1kq, \qquad (46)$$

and where L has been previously defined in (35).

This rather involved expression, which may be written in the form

$$\sigma_s^2=Be^{-\lambda s}\cos(\theta s+\omega)+Ce^{-\mu s}, \qquad (47)$$

is given for the sake of completeness. From it the values of the correlations r_s may be obtained directly. However, in nature the disturbances operating on an oscillating system can hardly be regarded as discrete impulses: in general the disturbances will be in the nature of a continuous fluctuating force acting on the system, and our next step will be to consider this case. Suppose, *e. g.*, that a stream of turbulent air flows at right-angles to the line of motion of the pendulum, then the nature of the problem remains essentially the same, with the exception that the discrete impulses imparted by the pellets is replaced by the continuous fluctuating force exerted by the turbulent component of the stream of air. This case of a continuous fluctuating force must be regarded as being of far more practical importance than the case just treated.

6. *The Limiting Case of a Continuous Fluctuating Disturbance.*

In the limiting case, when the time interval h approaches zero, we can develop the different functions in power series of h, retaining only such terms as ultimately prove to be significant. Thus, putting

$$\lambda=l+m/M, \qquad (48)$$

$$\beta^2=\alpha^2+l^2-\lambda^2, \qquad (49)$$

$$\beta_1^2=\beta^2+\lambda^2=\alpha^2+l^2, \qquad (50)$$

we find that

$$\left.\begin{aligned} &q=e^{-\lambda h}=1-\lambda h+\tfrac{1}{2}\lambda^2h^2,\\ &\theta=\beta h,\\ &k=2-h^2\beta^2,\\ &L=4h^2(\beta^2+\lambda^2)=4h^2\beta_1^2,\\ &\sigma^2=4\eta^2h^2 \end{aligned}\right\} \qquad (51)$$

In proceeding to the limit, we also find that (37) and (38) reduce to

$$\chi_s=\frac{e^{-\lambda sh}\sin(\beta sh+\gamma)}{4\lambda\beta\beta_1h^3}, \qquad (52)$$

where
$$\sin\gamma-\beta/\beta_1. \qquad (53)$$

If we now denote the time-lag sh by x, h may be regarded as the increment dx, and it follows from (51) and (52) that

$$\sigma^2\chi_x=\frac{\eta^2e^{-\lambda x}\sin(\beta x+\gamma)}{\alpha\beta\beta_1}dx. \qquad (54)$$

We may now apply this result to (39), replacing the summation by integral signs. However, it is advisable first to re-write (39) and adjust the limits of integration thus :

$$\begin{aligned}\frac{\sigma_\xi^2\,dx}{\sigma^2}&=\int_0^\infty(R_x\chi_{x+\xi}+R_{x+\xi}\chi_x)\,dx+\int_0^\xi R_x\chi_{\xi-x}\,dx\\&=\int_0^\infty(R_{x+\xi}+R_{x-\xi})\chi_x\,dx+\int_0^\xi(R_{\xi-x}-R_{x-\xi})\chi_x\,dx. \qquad (55)\end{aligned}$$

Hence, by virtue of (54)

$$\sigma_0^2=\frac{2\eta^2}{\lambda\beta\beta_1}\int_0^\infty R_xe^{-\lambda x}\sin(\beta x+\gamma)\,dx \qquad (56)$$

and

$$\begin{aligned}r_\xi=\frac{1}{2\int_0^\infty R_xe^{-\lambda x}\sin(\beta x+\gamma)}\Big[&\int_0^\infty(R_{x+\xi}+R_{x-\xi})e^{-\lambda x}\sin(\beta x+\gamma)\,dx\\&+\int_0^\xi(R_{\xi-x}-R_{x-\xi})e^{-\lambda x}\sin(\beta x+\gamma)\,dx\Big]. \qquad (57)\end{aligned}$$

This equation constitutes the general solution of our problem, and it must be regarded as an equation of far-reaching importance, for therein the parameters defining the statistical properties of the disturbed motion are described in terms of the parameters η^2 and R of the disturbing agency. If R is defined in any way whatsoever, the corresponding values of σ_0 and r can be evaluated either by ordinary or by numerical integration.

Suppose, *e. g.*, that R is a simple harmonic function $\cos b\xi$, which means that the disturbing force is also of the simple harmonic type. Here we have the well-known case of forced simple harmonic motion in which the pendulum eventually swings with the same period as that of the disturbing force, and hence the value of r should also be $\cos b\xi$. If, now, $R_\xi=\cos b\xi$ be substituted in (57), we do indeed find this to be the case.

In the *particular case* where R is an exponential function of the time-lag, $R_x=e^{-\mu x}$, the integration can be performed directly, and leads to the result

$$\sigma_0^2=\frac{2\eta^2(\mu+2\lambda)}{\lambda\beta_1^2\{\beta^2+(\mu+\lambda)^2\}} \qquad (58)$$

and

$$r_\xi=Be^{-\lambda\xi}\cos(\beta\xi+\omega)+Ce^{-\mu\xi}, \qquad (59)$$

where

$$\mathrm{B}=\frac{\mu\beta_1}{\beta(\mu+2\lambda)}\left\{\frac{\beta^2+(\mu+\lambda)^2}{\beta^2+(\mu-\lambda)^2}\right\}^{\frac{1}{2}}, \qquad (60)$$

$$\mathrm{C}=\frac{2\lambda\beta_1^{\,2}}{(\mu+2\lambda)\{\beta^2+(\mu-\lambda)^2\}}, \qquad (61)$$

and

$$\cos\omega=\frac{\beta(\beta^2+\mu^2-3\lambda^2)}{\beta_1\{(\beta^2+\mu^2+\lambda^2)^2-4\mu^2\lambda^2\}^{\frac{1}{2}}}. \qquad (62)$$

When we consider the velocities u_r instead of the displacements y_r of the pendulum, it is clear that the auto-correlation coefficient remains the same. From (15) it is readily seen that, in the limit, when h approaches zero,

$$\Sigma\delta_r^{\,2}=\mu^2\Sigma\epsilon_r^{\,2},$$

and hence the standard deviation of the velocity is given by

$$\frac{1}{n}\Sigma u_r^{\,2}=\mu^2\sigma_0^{\,2}. \qquad (63)$$

From (58) it is evident that the standard deviation σ_0 becomes infinite when $\lambda=0$, or when $q=1$, but this only happens when the pendulum is undamped, *i. e.*, when $l=0$ and when m/M is infinitely small. In Yule's original work he devotes special attention to this case, but apparently he did not realize that the amplitude of the pendulum then becomes infinitely large.

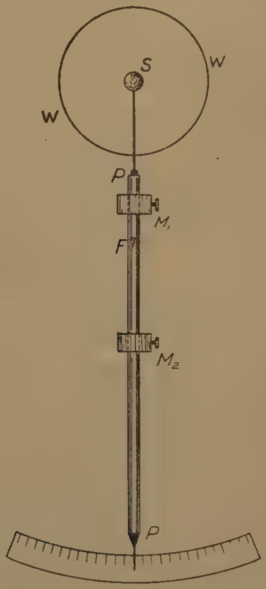
The practical application of the results obtained to experimental time series, *e. g.*, Wolfer's sunspot numbers, must be deferred to a subsequent publication, but there is another practical application which merits discussion. The relations established in (56) and (57) naturally lead to the question whether it is not possible to use a pendulum for the experimental investigation of the turbulent flow of a fluid. If the motion of a pendulum, subject to the turbulent force exerted on it by a fluid, be observed, then these equations should enable us to deduce the turbulent flow of the fluid itself, and the next paragraph will be devoted to the consideration of this question.

7. *The Experimental Determination of Turbulence.*

In the theory of turbulence, G. I. Taylor (1921) introduced the important concept of R_ξ, which he defines as the correlation coefficient of the turbulent velocity of a fluid at a given point and its velocity at the same point after a time ξ.

Now, for the further development of the theory of turbulence, it is of the greatest importance to determine R_ξ experimentally, but its determination by means of a hot-wire anemometer (Taylor, 1935) presents fairly serious experimental difficulties, and a relatively simple mechanical appliance must be considered as a very welcome addition to the equipment of the experimentalist, who concerns himself with the practical study of turbulence.

The proposed apparatus consists essentially of a rigid pendulum, as shown schematically in the accompanying figure. The straight rod PP is suspended at F, its mode of suspension being such that it is constrained to move in the plane of the paper. A thin steel needle projects upwards through a narrow slit in the cylindrical wind-tunnel WW and terminates in the sphere S. M_1 and M_2 are movable weights by which the moment of inertia and the period of the pendulum can be adjusted at will.



Pendulum for measuring the turbulence of an air-stream.

Furthermore, vanes (not shown in the diagram) can be attached to the pendulum in order to regulate the damping, and if a high degree of damping is desired, one of the vanes may be made to move within a fluid. A pointer at the lower-end of the pendulum moves over a suitable scale, and the position of the pendulum can thus be read off at any instant of time. Although the observation can be done by eye, it would, in general, be advisable to employ an automatic recording apparatus, such as a movie-camera, which would give a permanent record of the motion of the pendulum.

When a stream of air is allowed to pass through the wind-tunnel the sphere S is subjected to the turbulent force of the air, and the pendulum will start moving irregularly, and the departure y of the pointer from its zero position is recorded at small equal intervals of time. Having determined $y_1 \ldots . y_n$ for a large number of time-intervals, it is now possible to calculate r_ξ for different values of ξ.

Various ways of using this instrument for investigating the turbulent flow of a fluid suggest themselves, but for the present we shall confine our attention exclusively to the experimental determination of R_ξ, since this is a matter of such paramount importance in the theory of turbulence.

Having determined r_ξ as a function of ξ experimentally, we must now deduce the value of R_ξ. For this purpose we have available (57), which, however, is an integral equation and thus not directly adaptable for our purpose. It is possible, however, to derive an equation in a more suitable form by differentiating (57) directly, but before doing this it is advisable once more to adjust the limits of integration and write (57) in the form

$$r_\xi = \frac{1}{2\int_0^\infty R_x \chi_x \, dx}\left[\int_0^\infty R_x(\chi_{x+\xi}+\chi_{x-\xi})\, dx + \int_0^\infty R_x(\chi_{\xi-x}-\chi_{x-\xi})\, dx\right].$$

If the differentiation of r_ξ is carried out four times in succession, and the necessary eliminations are effected between the four equations thus obtained and the original equation, one readily arrives at the following expression :

$$\frac{4\eta^2}{\sigma_0^{\,2}} R_\xi = \beta_1^{\,4} r + 2(\beta_1^{\,2} - 2\lambda^2)\frac{d^2 r}{d\xi^2} + \frac{d^4 r}{d\xi^4} . \quad . \quad . \quad . \quad . \quad (64)$$

The same result may also be obtained by making use of equation (10),

$$\epsilon_r = y_r - kg\, y_{r-1} + q^2 y_{r\quad 2},$$

and computing

$$\frac{1}{n}\Sigma \epsilon_r \epsilon_{r+s},$$

which is equal to $\sigma^2 R_s$, in the limiting case when h approaches zero.

Since $R_\xi = r_\xi = 1$ when $\xi = 0$, equation (64) may also be written in the form

$$R_\xi = \frac{\beta_1^{\,4} r + 2(\beta_1^{\,2} - 2\lambda^2)\dfrac{d^2 r}{d\xi^2} + \dfrac{d^4 r}{d\xi^4}}{\beta_1^{\,4} + 2(\beta_1^{\,2} - 2\lambda^2)\left(\dfrac{d^2 r}{d\xi^2}\right)_0 + \left(\dfrac{d^4 r}{d\xi^4}\right)_0} . \quad . \quad . \quad . \quad . \quad (65)$$

Since r has been determined experimentally, its second and fourth derivatives with respect to ξ can also be computed, and in the above expression it remains to be seen whether β_1 and λ can also be determined experimentally. Both the period and the logarithmic decrement can be found in the ordinary way, and hence also β_1 according to (50). On the other hand, $\lambda=l+m/\mathrm{M}$ in the case of a simple pendulum, but as we are now dealing with a rigid pendulum, the moment of inertia enters into the original equation instead of the mass M, and accordingly we should write

$$\lambda=l+\mathrm{C}/\mathrm{I}, \qquad \ldots\ldots\ldots \qquad (66)$$

in which I is the moment of inertia of the pendulum about its axis of rotation and C is a constant. If this value be substituted for λ in (65), the latter may be regarded as containing the two unknowns R_ξ and C, since I can also be found experimentally. It thus becomes necessary to perform at least two experiments, keeping the turbulent flow of the fluid unaltered, but varying the moment of inertia of the pendulum.

Once the value of C has been found, equation (65) allows us to determine R_ξ for all values of ξ, and thus possibly to express it in analytic form.

Experimental work on these lines is proceeding at present, and the results will be reported at a later stage.

Summary.

With Yule's method of treating "disturbed" simple harmonic motion as a starting point, there is here developed the theory of the motion of a pendulum which is subject to a continuous fluctuating force. The theory leads to expressions whereby the statistical characteristics of the motion of the pendulum are expressed in terms of the statistical characteristics of the disturbing force.

A method is indicated whereby the character of the turbulent motion of a fluid can be experimentally determined by means of a suitably constructed rigid pendulum.

References.

G. U. Yule, Phil. Trans. Roy. Soc. A, vol. ccxxvi. p. 267 (1927).
G. T. Walker, Proc. Roy. Soc. Series A, vol. cxxxi. p. 518 (1931).
T. E. W. Schumann, Phil Mag. (1941).
G. I. Taylor, Proc. London Math. Soc. vol. xx. p. 196 (1921).
G. I. Taylor, Proc. Roy. Soc. Series A, vol. cli. p. 421 (1935).

XVII. *The Comma in Music*—II.

By A. F. Dufton, M.A., D.I.C.*

[Received December 9, 1941.]

So from old chronicles, where sleep in dust
Names that once filled the world with trumpet tones,
I build this verse.—H. W. Longfellow.

1. I write in praise of Major-General Thomas Perronet Thompson (1783–1869), Member of Parliament, Fellow of the Royal Society, and amateur of music.

When I penned the brief note "The Comma in Music," which appeared in this journal last September, I did so in simple ignorance of the work of General Thompson. It is now my pleasant duty, therefore, to invite attention to the importance of his contributions to the science and to the art of music.

2. It was pointed out by Perronet Thompson that if we desire to produce a scale in almost precisely just intonation, which will allow of an indefinite power of modulation, we can effect our purpose by the division of the octave into fifty-three equal parts. The ear cannot distinguish between this scale and the just, and in its practical application it admits of unlimited modulation in what is equal to exact intonation.

3. Thompson was insistent that temperament is a barbarous invention for saving trouble by playing out of tune ;—for playing in many keys, by playing in no key at all ;—for trying how much discordance the ear can be induced to bear, instead of how much harmony it can be accustomed to demand. "If men have not musical instruments that will play in tune in a variety of keys," he wrote, "they should be content to play in one, or else improve their instruments. But they ought not to play out of tune and call it music."

In 1829, in 'Instructions to my daughter for playing on the Enharmonic Guitar,' he proposed an important experiment designed to put his views to a practical test. It is of particular interest that the observations required in this test were made, thirty-five years later, by two men whose interest in musical acoustics was profound—Alexander J. Ellis, F.R.S., and Professor Dr. Hermann Helmholtz.

---

* Communicated by the Author.

"As a fair trial," Thompson wrote, "it may be hoped at some time to hear a church organ which shall produce correct harmony in a limited number of keys (on some principle like that proposed further on, in § 255),—played by an organist who will make the best practicable arrangement of church music in the supposed keys, by taking all the advantage possible of the power of change of key which he possesses,—and accompanied by a choir trained in the perception of correct harmony. The result would go far towards deciding, whether mankind at large really prefer the sweet water or the salt."

On the 20th of April *, 1864, Alexander Ellis took Professor Helmholtz to the Sunday School Chapel at 10 Jewin Street, Aldersgate, to hear the Enharmonic Organ, which was designed by General Thompson to play in just intonation in a variety of keys. In his 'Sensations of Tone,' which was translated by Ellis, Helmholtz wrote :—

"In London I had also an opportunity of comparing the intonation of this instrument with the natural intonation of singers who had learned to sing without any instrumental accompaniment at all, and are accustomed to follow their ear alone. This was the Society of Tonic Sol-fa-ists, who are spread in great numbers (there were 150,000 in 1862) over the large cities of England, and whose great progress is of much importance for the theory of music.

. . . .

"After this experience, I think no doubt can remain, if ever any doubt existed, that *the intervals which have been theoretically determined in the preceding pages, and there called natural, are really natural for uncorrupted ears ; that moreover the deviations of tempered intonation are really perceptible and unpleasant to uncorrupted ears ; and lastly that, notwithstanding the delicate distinctions in particular intervals, correct singing by natural intervals is much easier than singing in tempered intonation.*

. . . .

"Those who have no time for such observations, should at any rate glance over the literature of the period during which equal temperament was introduced. When the organ took the lead among musical instruments it was not yet tempered. And the pianoforte is doubtless a very useful instrument for making the acquaintance of musical literature, or for domestic amusement, or for accompanying singers. But for artistic purposes its importance is not such as to require its mechanism to be made the basis of the whole system of music."

---

* On April 14th, Helmholtz delivered the CROONIAN LECTURE to the Royal Society.

4. Helmholtz's reference to the leading place taken by the organ lends point to an appreciation expressed by an earlier amateur of music, Samuel Pepys :—

"1667 April 21st (Lord's Day)—To Hackney Church, where very full, and found much difficulty to get pews, I offering the sexton money and he could not help me. So my wife and Mercer ventured into a pew, and I into another. A knight and his lady very civil to me when they came, being Sir G. Viner and his lady, rich in jewells, but most in beauty—almost the finest woman I ever saw. That which I went chiefly to see was the young ladies of the schools, whereof there is a great store, very pretty ; and also the organ, which is handsome, and tunes the psalm and plays with the people, which is mighty pretty, and makes me mighty earnest to have a pair at our church (St. Olaf, Hart-street), I having almost a mind to give them a pair, if they would settle a maintenance on them for it."

5. The important contribution made by Perronet Thompson to English singing was acknowledged by the Rev. John Curwen, who in 1853 founded the Tonic Sol-fa Association. Curwen adopted Thompson's view that for the cultivation of correct harmony just intonation is essential. "What seems to be wanted both by singers and violinists," wrote Thompson in 1829, "is the habit of viewing notes simply by reference to their place in the octave, or their relation to the existing Key-note; instead of attempting to refer them to a scale of immovable sounds. And for this purpose it would only be necessary to employ the syllables *do, re, mi, fa, sol, la, si,* according to what was probably their original intention—by taking *do* for the Key-note in every key, and the rest for Second, Third, &c. in succession ; which is what the French call *solfier par transposition*. The Minor sounds of *re, mi, la, si* might be expressed by changing the final vowel into the Italian *u*."

After hearing the singing of children in the British and Foreign School behind the chapel in Tottenham Court Road, Professor Helmholtz wrote to Mr. Curwen to say, "I think that what I saw shewed the complete success of your system, and I was peculiarly interested by it, because during my researches in musical acoustics I came from theoretical reasons to the conviction that this was the natural way of learning music, but I did not know that it had been carried out in England with such beautiful results." And, as he remarked to Mr. Ellis, "We could not do that in Germany."

6. In view of General Thompson's suggestion that we should improve our instruments, it is fitting that we should contemplate the future which lies before us. It may well be that the recent advances in physics

will enable us to safeguard the principles upon which musical art is founded and to avert the threat that the mechanism of our present instruments and attention to their convenience will lord it over the natural requirements of the ear.

For church music, it is not so important now that there should be an enharmonic organ in every church or chapel : the immediate need is rather for a good arrangement of church music, in just intonation, recorded so as to be available for reproduction by means of a loud-speaker system. The recording of music in the form of gramophone records and as sound-films marks an important advance, and it is of some interest that records can now be made of an instrumentalist playing all three parts of a trio, or of a singer taking both parts of a duet.

The basis of instrument design has been developed to an extraordinary degree : if we now wish, for example, to build an organ which will speak like a violin, not only can we faithfully reproduce the finest note which the best violin can utter ; not only can we transpose this note so that the same perfection obtains for every note in the gamut, but we can also design our organ so that it is always in perfect tune.

May we not envisage an establishment, directed by competent musicians and equipped with all that science has to offer, which, after the necessary research and development, will not only prepare and make available a superlative version of all that is best in music, but which will also provide ambrosial music, the like of which has not been heard since Orpheus touched upon his lute ?

The King's Lodge,
Hunton Bridge, Hertfordshire.

XVIII. *The Comma in Music*—III.

By A. F. Dufton, M.A., D.I.C.*

[Received December 22, 1941.]

No man that wadeth in learning or contemplation throughly, but will find that printed in his heart Nil novi super terram.—F. Bacon.

1. Mr. Ll. S. Lloyd, whose recent books on 'Music and Sound' and 'The Musical Ear' have contributed so much to the understanding of the physical aspects of music and musical appreciation, has kindly invited my attention to the division of the octave into 53 equal parts by Nicholas Mercator in the seventeenth century, citing as a reference the second English edition of Helmholtz's 'Tonempfindungen.'

On consulting the twentieth appendix of this work, which was added by the translator, I was surprised to find that a cycle of 53, with a fifth of 31 and a major third of 18 degrees, is ascribed to Mercator. This major third is a comma sharp and, as a temperament, the cycle is hardly worthy of notice.

In the absence of any authentic reference, and as it seemed unlikely that a capable mathematician like Mercator would have published such an unsatisfactory scheme, I decided not to mention this cycle in the note which I prepared about the work of General Perronet Thompson.

2. After a vain search for any account by Mercator of an harmonic cycle, I have now discovered that, in all probability, Mercator did not publish his invention. An interesting account of it was given, however, in 'A Treatise of The Natural Grounds, and Principles of Harmony,' by William Holder, D.D., F.R.S. This work was published in 1694, and the following is an extract of the relevant passage, taken from the edition printed in 1731 :—

3. "*MERSENNUS* finds, by his Calculation, 58¼ Comma's, and somewhat more, in an *Octave*: But the late *Nicholas Mercator*, a Modest Person, and a Learned and Judicious Mathematician, in a Manuscript of his, of which I have had a Sight, makes this Remark upon it ; *In solvendo hoc Problemate aberrat Mersennus* : And he, working by the Logarithms, finds out but 55, and a little more ; and from thence has deduced an

---

* Communicated by the Author.

ingenious Invention of finding and applying a least Common Measure to all Harmonic Intervals, not precisely perfect, but very near it.

"Supposing a Comma to be $\frac{1}{53}$ part of Diapason; for better Accommodation rather than according to the true Partition $\frac{1}{55}$, which $\frac{1}{53}$ he calls an Artificial Comma, not exact, but differing from the true Natural Comma about $\frac{1}{20}$ part of a Comma, and $\frac{1}{1000}$ of Diapason (which is a Difference imperceptible) then the Intervals within Diapason will be measur'd by Comma's according to the following Table; which you may prove by adding two, or three, or more of these Numbers of Comma's, to see how they agree to constitute those Intervals, which they ought to make; and the like by substracting

Intervals	$\frac{0}{53}$	Intervals	$\frac{0}{53}$
Comma	1	4th	22
Diesis	2	*Tritone*	26
Semit. Minus	3	*Semidiapente*	27
Semit. Medium	4	5th	31
Semit. Majus	5	6th *Minor*	36
Semit. Maximus	6	6th *Major*	39
Tone Minor	8	7th *Minor*	45[*]
Tone Major	9	7th *Major*	48
3rd *Minor*	14	*Octave*	53
3rd *Major*	17		

"This I thought fit, on this Occasion, to impart to the Reader, having Leave so to do from Mr. *Mercator's* Friend, to whom he presented the said Manuscript."

4. Having now made amende for describing as a novelty a discovery which was known two centuries ago, I think fit, on this occasion, to crave the reader's indulgence to the author for his whilom failure "to collect that previous knowledge of all that has been already done with the same view, which . . . he ought to acquire . . . before he calls on the world at large to participate in his improvements and discoveries"[†].

The King's Lodge,
Hunton Bridge, Hertfordshire.

---

[*] This number corresponds to the frequency-ratio 9/5, the Acute Minor 7th.

[†] Young, T., 'A Course of Lectures on Natural Philosophy and the Mechanical Arts,' London, 1807.

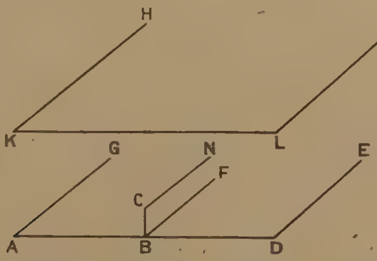
XIX. *Heat Flow when the Boundary Condition is Newton's Law.*

By J. H. Awbery, B.A., B.Sc., F.Inst.P.*

[Received November 27, 1941.]

§ 1. *Introduction.*

In a paper published in 1929 (Phil. Mag. vii. p. 1143) under the above title, I pointed out that solutions to heat-flow problems in which the boundary condition is Newton's law may under certain circumstances be obtained by adding further terms to a solution in which the same boundaries are isothermals. I gave two examples, and it has recently been pointed out by Schofield (Phil. Mag. xxxi. p. 471) that the second of them is erroneous.



Since he did not show how to correct it, and indeed implies that it and similar problems lie outside the scope of the method, and since a correspondent mentions that he also has been unable to solve the problem, I give here the corrected result.

§ 2. *Girder Projecting into Insulation.*

The problem which forms the subject of Schofield's criticism is shown in the figure. Here the plane ABD and the girder BC are supposed to be at a single temperature, taken as zero, and the plane KL is the one to which Newton's law applies. The solution when KL is at a constant temperature v_0 is easily obtained by conformal transformation, and is given by equation (VII.) of the 1929 paper. No objection is raised to this part of the solution, but it is to be noted that it is an implicit equation in which the x-coordinate occurs only as $\cosh \pi x/l$ and $\sinh^2 \pi x/l$, so that changing the sign of x does not alter the equation.

---

* Communicated by the Author.

Thus, although the solution was actually obtained by considering only the semi-infinite plane to the right of NCBF, it gives a symmetrical solution applicable also to the left half of the plane.

To pass from the solution in which KL is isothermal to that in which Newton's law applies, a term

$$v_2 = \mathrm{A}\int_0^\infty \sinh my \sin mx \, dx$$

was added to the temperature as previously calculated. This quantity, whatever the finite value of A, vanishes when $x=0$ or when $y=0$.

It is a solution of the fundamental differential equation, and is not infinite for any value of x or of y within the region covered by the problem. It thus satisfies all the necessary conditions over ABD and BC; the fact that it does not of itself make the prolongation of BC into a flow-line is irrelevant if that prolongation is not part of the boundary. All that remains is to secure that Newton's law shall hold for the combined expression (v_1+v_2) over the plane KL, and not merely, as in the 1929 paper, over the half-plane to the right of BC.

The analysis given in the 1929 paper holds as far as the equation in the middle of p. 1152:

$$-\mathrm{A}\int_0^\infty \sin mx \, (\mathrm{E} \sinh ml_1 + \mathrm{K}m \cosh ml_1) dm = \frac{\mathrm{K}v_0(\alpha+\beta) \sinh \pi x/l_1}{\mu(x)}.$$

(There is a misprint in the original, where $\sinh \pi v/l_1$ occurs for $\sinh \pi x/l_1$).

The right-hand side has now to be expanded in a Fourier's integral, valid for $-\infty < x < \infty$; in the original paper an expansion valid for $0 \leqslant x < \infty$ was given, which would have been applicable if $\mu(x)$, like $\sin mx$, had been an odd function. To obtain the correct expansion involves writing

$$\int_{-\infty}^\infty \quad \text{for} \quad 2\int_0^\infty$$

and replacing $\sin m\lambda \sin mx$ by $\cos m(\lambda - x)$.

Thus, the last equation in the paper becomes

$$v_2 = -\frac{\mathrm{K}v_0(\alpha+\beta)}{\pi}\int_0^\infty dm \int_{-\infty}^\infty \frac{\sinh \pi\lambda/l_1 \cos m(\lambda-x) \sinh my \, d\lambda}{\mu(\lambda)(\mathrm{E} \sinh ml_1 + \mathrm{K}m \cosh ml_1)},$$

with the same notation as before; and, with this modification, the solution becomes complete and accurate.

XX. *The Fermi-Dirac Functions.*

By F. C. Auluck, Dyal Singh College, Lahore *.

[Received May 8, 1941.]

1. In recent years various applications (connected with the Fermi-Dirac statistics) of the Functions

$$F(t)=\int_0^\infty \frac{\phi'(x)\,dx}{e^{x-t}+1}$$

have been made to physical and astrophysical problems. J. L. B. Cooper † has recently given a new method for evaluating these functions when $\phi'(x)=x^k$. In this note I have worked out the corresponding case for any function $\phi(x)$.

2. If

$$F(t)=\int_0^\infty \frac{\phi'(x)}{e^{x-t}+1}\,dx \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and $w=u+iv$, where $0<v<1$, we get

$$\int_{-\infty}^\infty e^{-wt}F(t)\,dt=\frac{\pi}{\sin \pi w}\int_0^\infty e^{-wx}\phi'(x)\,dx, \quad . \quad . \quad . \quad . \quad (2)$$

and so, by the inversion formula for the Laplace Transform,

$$F(t)=\frac{1}{2i}\int_{a-i\infty}^{a+i\infty}\frac{e^{wt}}{\sin \pi w}\Phi(w)\,dw, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where $$0<\alpha<1,$$

and $$\Phi(w)=\int_0^\infty e^{-wx}\phi'(x)\,dx.$$

Since $$\frac{\pi w}{\sin \pi w}=1+2\sum_{n=1}^{\infty} C_n w^{2n},$$

where $$C_n=(2^{n-1}-1)B_n\frac{\pi^{2n}}{(2n)!},$$

we can write the integral (3) as

$$F(t)=\frac{1}{2i\pi}\int_{a-i\infty}^{a+i\infty}\left(\frac{1}{w}+2\sum_{m=1}^{n} C_m w^{2m-1}\right)e^{wt}\Phi(w)\,dw$$
$$+\frac{1}{2i\pi}\int_{a-i\infty}^{\alpha+i\infty}\psi(w)e^{wt}\,dw, \quad . \quad . \quad (4)$$

---

* Communicated by Dr. D. S. Kothari.

† J. L. B. Cooper, Phil. Mag. pp. 187–189 (Sept. 1940).

where $\psi(w)$ denotes the expression

$$2 \sum_{m=n+1}^{\infty} \mathrm{C}_m w^{2m-1} . \Phi(w).$$

The asymptotic expansion of $\mathrm{F}(t)$ $(t>0)$ can be found from the expansion of its transform near $w=0$.

If $\phi(x)=0(x^k)$ as $x \rightarrow \infty$, then, as $w \rightarrow 0$,

$$\psi(w)=0(w^{2n-k+1}).$$

It can be shown * that

$$\frac{1}{2i\pi}\int_{\alpha-i\infty}^{\alpha+i\infty} \psi(w) e^{wt}\, dw=0(t^{-2n+k-2}).$$

It is sufficient to evaluate the integral

$$f(t)=\frac{1}{2i\pi}\int_{\alpha-i\infty}^{\alpha+i\infty} \frac{e^{wt}}{w}\, \Phi(w)\, dw, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

for the other terms in (4) can be evaluated by differentiating $f(t)$ under the integral sign. Applying the inversion formula of Laplace Transform again to (5), we get

$$\frac{1}{w}\,\Phi(w)=\int_0^{\infty} e^{-wt} f(t)\, dt.$$

But

$$\frac{1}{w}\,\Phi(w)=\frac{1}{w}\int_0^{\infty} e^{-wx} \phi'x)\, dx,$$

$$=\int_0^{\infty} e^{-wx} \phi(x)\, dx,$$

for the function $\phi(x)$ can always be taken to satisfy the condition $\phi(0)=0$, since we are only given $\phi'(x)$. Hence

$$\int_0^{\infty} e^{-wt}\{f(t)-\phi(t)\} dt=0,$$

whatever w may be. Therefore, by a theorem of Lerch †, $f(t)=\phi(t)$. So we get

$$\mathrm{F}(t) \approx \phi(t)+2 \sum_{n=1}^{\infty} \mathrm{C}_n \phi^{2n}(t).$$

---

* *Vide* J. L. B. Cooper, *loc. cit.*

† Lerch, *Acta Math.* xxvii. pp. 339–351 (1903).

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

XXI. *Corresponding Problems in Periodic and Steady Flow.*

By GEORGE GREEN, D.Sc., Lecturer in Applied Physics
in the University of Glasgow *.

[Received November 17, 1941.]

PART II.

IN the problems relating to point sources and line sources, which we have discussed in Part I., the boundary conditions have been chosen merely to illustrate the application of the special method of obtaining continuous flow solutions explained above. The solutions obtained provide a group of analogous results in heat conduction, steady fluid motion, electrostatics, and other subjects to which the fundamental equation (3) is applicable.

By altering the boundary condition to

$$K_1\frac{dv_1}{dn_1}=K_2\frac{dv_2}{dn_2},$$

that governing transmission of heat across a surface separating two different media, we may obtain by the same method another group of analogous results. One fundamental problem involving this boundary condition has already been discussed in ref. (4), and on page 638 of this paper attention is drawn to the theorem stated in Part I., pp. 106, 107, which we have used to determine the particular terms in any periodic flow solution that provide a contribution to the corresponding continuous flow solution, in the application of the method of the present investigation. Other problems involving this boundary condition may be discussed later.

It is of greater interest at present to examine the application of this method to problems involving flow of fluid past an obstacle, or to problems in which the flow is along specified boundaries. The solution of the fundamental problems of flow past a cylinder and flow past a sphere may serve to illustrate the utility of the method in such cases. Solutions are also given for certain diffraction problems.

---

* Communicated by the Author.

§ 1. *Flow past a Cylinder.*

Periodic Flow Solution.—Consider first the case of periodic heat flow past an infinite cylinder of radius a. We may take as incident wave

$$v=\frac{q}{2\mathrm{K}i\lambda}e^{ikt-i\lambda x} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

This represents a plane wave train advancing in the direction of OX, the flow of heat per unit area of any plane perpendicular to OX being $\frac{q}{2}e^{ikt}$. The problem is—to determine how this flow is modified by the presence of a cylinder which is impermeable to heat. The solution to the proposed problem depends on the expansion

$$e^{-i\lambda x}=\mathrm{I}_0(i\lambda\rho)+2(-1)^n\sum_1^\infty \mathrm{I}_n(i\lambda\rho)\cos n\theta, \quad . \quad . \quad . \quad . \quad (53)$$

in which ρ is the radius vector drawn from the origin to any point whose co-ordinates are (ρ, θ) referred to the axis of the cylinder. By means of this expansion the incident wave (52) is replaced by a group of waves, each of which is incident radially on the cylinder; and the solution is found by adding to each incident wave of type $\mathrm{I}_n(i\lambda\rho)$ a reflected wave of type $\mathrm{K}_n(i\lambda\rho)$. Thus the required solution of the problem is represented by

$$v=\frac{qe^{ikt}}{2\mathrm{K}i\lambda}[\mathrm{I}_0(i\lambda\rho)+\mathrm{A}_0\mathrm{K}_0(i\lambda\rho)+2(-1)^n\sum_1^\infty \mathrm{T}_n\cos n\theta], \quad . \quad . \quad (54)$$

where $\mathrm{T}_n=\mathrm{I}_n(i\lambda\rho)+\mathrm{A}_n\mathrm{K}_n(i\lambda\rho)$, and where the constants, A_n, are to be chosen to secure the fulfilment of boundary conditions at the surface of the cylinder. Taking the boundary condition

$$\frac{dv}{d\rho}=0, \text{ at } \rho=a, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (55)$$

applicable to a cylinder which is impermeable to heat, we find that

$$\mathrm{A}_n=-\frac{\mathrm{I}'_n(i\lambda a)}{\mathrm{K}'_n(i\lambda a)}\cos n\theta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (56)$$

and with these values inserted in it (54) represents the periodic flow solution required.

Continuous Flow Solution.—The corresponding steady flow solution, representing flow past the cylinder, is readily obtained by means of former results. The periodic incident wave (52) yields

$$v_c=-\frac{q}{2\mathrm{K}}x \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (57)$$

as corresponding steady flow solution—a result obtained in (17′) of Part I.

In order to obtain the continuous flow solution corresponding to any term of the reflected wave-system, it is convenient first to determine the continuous flow solution corresponding to the periodic solution

$$v_p=\frac{e^{ikt}\mathrm{K}_n(i\lambda\rho)}{(i\lambda a)\mathrm{K}_n'(i\lambda a)}\cos n\theta. \qquad (58)$$

By a well-known property of Bessel functions we may write

$$\frac{1}{z\mathrm{K}_n'(z)}=\frac{\mathrm{K}_n(z)\mathrm{I}_n'(z)}{\mathrm{K}_n'(z)}-\mathrm{I}_n(z), \qquad (59)$$

and, with $z=i\lambda a$, this enables us to express v_p as the difference of two periodic solutions, for which the corresponding continuous flow solutions are indicated by results given in (51) and (21) of Part I. respectively. Thus the continuous flow solution corresponding to (58) is given by

$$v_c=\left\{-\frac{1}{n}\cdot a\cdot\frac{d}{da}\left\{\frac{1}{2n}\left(\frac{a}{\rho}\right)^n\right\}-\frac{1}{2n}\left(\frac{a}{\rho}\right)^n\right\}\cos n\theta. \qquad (60)$$

$$=-\frac{1}{n}\left(\frac{a}{\rho}\right)^n\cos n\theta. \qquad (60')$$

Considering now the successive terms of the reflected wave-system represented by the K_n terms in (54), each term is of the type $v_p.\mathrm{I}_n'(i\lambda a)$, where v_p is the periodic solution (58). In accordance with the method employed in Part I., we may expand $\mathrm{I}_n'(i\lambda a)$ in a series of ascending powers of λ^2, and apply the theorem of Part I., pp. 106, 107, to the successive terms. In following out this process we find that the continuous flow solution corresponding to the periodic solution term $n=0$ is zero, in accordance with equation (B). The periodic solution term $n=1$ contributes

$$v_c=-\frac{q}{2\mathrm{K}}\cdot\frac{a^2}{\rho}\cos\theta \qquad (61)$$

to the continuous flow solution. The terms $n=2$, 3, etc., give zero contributions in accordance with the same general theorem.

The steady flow solution obtained by adding (57) and (61) is thus

$$v=-\frac{q}{2\mathrm{K}}x-\frac{q}{2\mathrm{K}}\frac{a^2}{\rho}\cos\theta, \qquad (62)$$

which is in complete agreement with the analogous solution representing the steady flow of an incompressible fluid past a cylinder. In the fluid analogue V takes the place of $\frac{q}{2\mathrm{K}}$, V being the uniform velocity of the fluid at infinite distance parallel to the axis OX.

§ 2. *Flow past a Sphere.*

Periodic Flow Solution.—The investigation in the case of the sphere is identical with that in the case of the cylinder. We again take (52) as the incident wave: and in this case the solution depends on the expansion

$$e^{-i\lambda x}=2^{\frac{1}{2}}\pi^{\frac{1}{2}}\sum_{0}^{\infty}(-1)^n(n+\tfrac{1}{2})\frac{I_{n+\frac{1}{2}}(i\lambda r)}{(i\lambda r)^{\frac{1}{2}}}P_n(\cos\theta). \quad . \quad . \quad . \quad (63)$$

By using this expansion in (52) above we convert the plane incident wave into a group of waves each of which is incident radially on the sphere, and the solution is found by adding to each incident wave of type $I_{n+\frac{1}{2}}$ a reflected wave of type $K_{n+\frac{1}{2}}$. The periodic solution in any particular case is given by

$$v=\frac{q2^{\frac{1}{2}}\pi^{\frac{1}{2}}e^{ikt}}{2K}\sum_{0}^{\infty}(-1)^n(n+\tfrac{1}{2})\frac{T_n}{(i\lambda)^{\frac{3}{2}}r^{\frac{1}{2}}}P_n(\cos\theta), \quad . \quad . \quad . \quad . \quad (64)$$

where $T_n=I_{n+\frac{1}{2}}(i\lambda r)+A_nK_{n+\frac{1}{2}}(i\lambda r)$, and where the constants, A_n, are to be chosen to secure the fulfilment of a boundary condition at $r=a$, the surface of the sphere. Taking again the boundary condition

$$\frac{dv}{dr}=0, \text{ at } r=a, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (65)$$

we find that

$$A_n=-\frac{zI'(z)-\frac{1}{2}I(z)}{zK'(z)-\frac{1}{2}K(z)}=-\frac{[I]'(z)}{[K]'(z)}, \quad . \quad . \quad . \quad . \quad . \quad (66)$$

where K and I represent Bessel functions of order $n+\frac{1}{2}$, and z is written instead of $(i\lambda a)$. Expressions (64) and (66) together represent the periodic solution.

Continuous Flow Solution.—The periodic incident wave again yields

$$v=-\frac{q}{2K}x \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (57)$$

as corresponding steady flow solution.

In order to determine the steady flow solution corresponding to the reflected wave, we consider first the periodic solution

$$v_p=\frac{e^{ikt}K_{n+\frac{1}{2}}(i\lambda r)}{zK'(z)-\frac{1}{2}K(z)}P_n(\cos\theta). \quad . \quad . \quad . \quad . \quad . \quad . \quad (67)$$

The procedure followed is to transform the expression by means of the theorem relating to Bessel functions stated in (59) above. This theorem, to suit the present purpose, may be expressed in the form

$$\frac{1}{[K]'(z)}=\frac{K(z)[I]'(z)}{[K]'(z)}-I(z). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (68)$$

It allows us to express v_p as the sum of two periodic solutions, to each of which the method described in Part I., p. 108, can be applied. The continuous flow solution corresponding to (67) is thus found to be given by

$$v_c = -\frac{1}{(n+1)}\left(\frac{a}{r}\right)^{n+\frac{1}{2}} \mathrm{P}_n(\cos\theta). \quad . \quad . \quad . \quad . \quad . \quad . \quad (69)$$

Now the successive terms of the reflected wave-system, represented by the $\mathrm{K}_{n+\frac{1}{2}}(i\lambda r)$ terms in (64), are all of the form $v_p\frac{[\mathrm{I}]'(z)}{(i\lambda)^{3/2}}$, where v_p is the periodic solution (67). It is only necessary to expand $\frac{[\mathrm{I}]'(i\lambda a)}{(i\lambda)^{3/2}}$ in a series of ascending powers of λ^2, and apply the theorem of Part I. to the successive terms. From the well-known expression for $\mathrm{I}_{n+\frac{1}{2}}(i\lambda a)$, we obtain

$$\frac{[\mathrm{I}]'_{n+\frac{1}{2}}(i\lambda a)}{(i\lambda)^{3/2}} = \frac{n(i\lambda a)^{n-1}}{2^{n+\frac{1}{2}}\pi(n+\frac{1}{2})} + \text{higher powers of } \lambda. \quad . \quad . \quad (70)$$

In accordance with the theorem referred to the only term of the reflected wave-system which provides a contribution to the continuous flow solution is that for which $n=1$, which yields, in virtue of (70),

$$v_c = -\frac{q}{4\mathrm{K}}\frac{a^3}{r^2}\cos\theta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (71)$$

The contribution from the $n=0$ term is zero, by the theorem as stated in equation B; and a like result holds for the $n=2$, 3, 4, etc., terms.

The complete steady flow solution is obtained by combining (57) and (71), and is given by

$$v = -\frac{q}{2\mathrm{K}}x - \frac{q}{4\mathrm{K}}\cdot\frac{a^3}{\rho^2}\cos\theta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (72)$$

To apply this result to the analogous solution representing flow of an incompressible fluid past a sphere, V being the uniform velocity of the fluid at infinite distance from the sphere in the direction OX, we have merely to replace $\frac{q}{2}$ by V and K by unity.

§ 3. *Diffraction Problems relating to Flow through an Aperture in a Thin Plane Rigid Wall.*

Circular Aperture of Radius a.—The solution to the problem—to determine the steady flow of fluid through a circular aperture in a plane rigid wall—is given in Lamb's 'Hydrodynamics,' 3rd edn. § 102, 3°. The object of the present investigation is to determine the solution to the corresponding periodic flow problem—by employing the relations established above between periodic and continuous flow solutions of (1) and (3) respectively (Part I.).

Continuous Flow Problem.—With the centre of the aperture as origin, with the Z-axis perpendicular to the plane of the thin plane screen or partition, the co-ordinates of any point being (ρ, z); the solution of the steady flow problem requires the determination of a velocity potential function, $\phi(\rho, z)$, to satisfy Laplace's equation, (3) of Part I., and the boundary conditions, at $z=0$,

$$\phi=\text{constant over the aperture,}\quad \rho<a, \quad . \quad . \quad . \quad . \qquad (73)$$

$$\frac{d\phi}{dz}=0, \text{ over the screen,}\quad \rho>a. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (74)$$

The problem is identical with that of determining the electrostatic potential at any point in space due to the presence of an electrified disk, occupying the position of the aperture, and charged to a constant potential. In the analogous heat conduction problem (6), ϕ is to be identified with temperature at any point in an infinite medium surrounding a flat circular disk which is kept at a constant temperature.

The solution given by Lamb to the continuous flow problem is

$$v=\frac{2}{\pi}\theta_0\int_0^\infty dk\,e^{-kz}J_0(k\rho)\frac{\sin ka}{k}. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (75)$$

This expression, which is adapted to suit the heat conduction analogue, gives $v=\theta_0$ as the constant temperature on the disk $\rho<a$, and zero gradient perpendicular to the screen.

To suit our purpose the solution (75) may be expressed by infinite series—one for the region $r<a$, and the other for the region $r>a$. By expanding $\sin ka$, and by using the well-known relation

$$\int_0^\infty dk\,e^{-kz}J_0(k\rho)=\frac{1}{r}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (76)$$

we obtain the expansion

$$v=\frac{2}{\pi}\theta_0\left\{\frac{a}{r}-\frac{a^3}{3\,!}\frac{\partial^2}{\partial z^2}\left(\frac{1}{r}\right)+\frac{a^5}{5\,!}\frac{\partial^4}{\partial z^4}\left(\frac{1}{r}\right)-\text{, etc.}\right\}, \quad . \quad . \quad . \quad . \qquad (77)$$

applicable to the region $r>a$. Now the general solution to Laplace's equation must be of the form

$$v=\theta_0\left\{\frac{A_0}{r}P_0(\cos\theta)+\frac{A_1}{r^2}P_1(\cos\theta)+\frac{A_2}{r^3}P_2(\cos\theta)+\text{, etc}\right\}, \quad . \qquad (78)$$

where $\cos\theta=\frac{z}{r}$, the axis OZ being an axis of symmetry with reference to v. In this case too, owing to the condition (74), the odd terms containing A_1, A_3, A_5, etc. must disappear. We can therefore express the solution of the continuous flow problem, for region $r>a$, in the form (78), the even coefficients A_0, A_2, A_4, etc. being found by equating corresponding terms in (78) and (77). The values of the required coefficients are given by

$$A_0=\frac{2}{\pi}a : A_2=-\frac{2}{\pi}\frac{a^3}{3} : A_4=\frac{2}{\pi}\frac{a^5}{5} : \text{and so on.} \quad . \quad . \quad . \quad . \quad (79)$$

It is to be noted that the above solution yields

$$v=\frac{2}{\pi}\theta_0 \sin^{-1}\left(\frac{a}{\rho}\right) \ldots \{z=0, \rho>a\}, \quad . \quad . \quad . \quad . \quad . \quad (80)$$

as the temperature distribution over the screen. In view of the method to be used below to solve the corresponding periodic flow problem, we remark that if temperature conditions over the $z=0$ plane

$$v=\theta_0, \ \rho<a : v=\frac{2}{\pi}\theta_0 \sin^{-1}\frac{a}{\rho}, \ \rho>a \quad . \quad . \quad . \quad . \quad . \quad (81)$$

were given, instead of the two conditions (73) and (74) above, the solution in both cases would be the same. Taking the two Fourier integral solutions of Laplace's equation

$$v=\theta_0\int_0^\infty dk e^{-kz}J_0(k\rho)\int_0^a d\rho'\rho' J_0(k\rho'), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (82)$$

$$v=\theta_0\int_0^\infty dk e^{-kz}J_0(k\rho)\int_a^\infty d\rho'\rho' J_0(k\rho') \, . \frac{2}{\pi}\sin^{-1}\left(\frac{a}{\rho}\right), \quad . \quad . \quad (83)$$

the first imposes the condition $v=\theta_0$, $\rho<a$: the second imposes the condition $v=\frac{2}{\pi}\theta_0 \sin^{-1}\frac{a}{\rho}$, $\rho>a$: on plane $z=0$. It is easy to verify that the sum of these solutions is given by (75) above ((6), pp. 943–5), and it follows that (80) represents that temperature distribution on the screen which is required to annul the normal gradient on the screen, in the presence of a disk kept at steady temperature θ_0.

Periodic Flow Problem.—Considering in particular the heat conduction analogue of the periodic flow problem, the conditions to be satisfied are now

$$v=\theta_0 e^{ik't} \ldots \text{over the disk } \rho<a, \quad . \quad . \quad . \quad . \quad . \quad (84)$$

$$\frac{dv}{dz}=0 \text{ at } z=0, \text{ over region } \rho>a, \quad . \quad . \quad . \quad . \quad . \quad (85)$$

and

$$(\nabla^2+\lambda^2)v=0, \text{ where } \lambda^2=-\frac{ik'}{\kappa}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (86)$$

in the notation used in Part I.

As in the case of the steady flow problem just considered, the required solution may be regarded as the sum of two solutions—the first of which imposes the condition (84), and the second of which imposes the condition

$$v=\theta_0 e^{ik't}F(\rho), \text{ over the screen, } \rho>a. \quad . \quad . \quad . \quad . \quad (87)$$

In order to solve the periodic flow problem, $F(\rho)$ has to be determined so that the sum of these two solutions satisfies the condition (85).

Since $F(\rho)$ refers to the region $\rho>a$, following upon the discussion given above of the constant flow problem, we take the general solution of Poisson's equation, subject to condition (85), in the form $(r>a)$

$$v=\theta_0 e^{ik't}\left\{\left(\frac{a}{r}\right)^{\frac{1}{2}}\sum_0^\infty A'_{2n}\frac{K_{2n+\frac{1}{2}}(i\lambda r)}{K_{2n+\frac{1}{2}}(i\lambda a)}P_{2n}(\cos\theta)\right\}. \quad . \quad . \quad . \quad (88)$$

Each term of this summation is itself a periodic solution of the wave-equation satisfying (85) ; A_0', A_2', etc., being arbitrary constants. The steady flow solution corresponding to each term can now be derived from it by the general process employed throughout this investigation. Considering the typical periodic solution

$$v_p=C.\frac{e^{ik't}K_{2n+\frac{1}{2}}(i\lambda r)}{K_{2n+\frac{1}{2}}(i\lambda a)}P_{2n}(\cos\theta), \quad . \quad . \quad . \quad . \quad . \quad (89)$$

in which C does not contain λ or t, and making use of the transformation

$$\frac{1}{K(z)}=zI'(z)-\frac{zK'(z)I(z)}{K(z)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (59)$$

we find by the process described in Part I.,

$$v_c=C.\left(\frac{a}{r}\right)^{2n+\frac{1}{2}}P_{2n}(\cos\theta). \quad . \quad . \quad . \quad . \quad . \quad . \quad (90)$$

The continuous flow solution corresponding to (88) is therefore represented by

$$v=\theta_0\left\{A_0'\frac{a}{r}P_0(\cos\theta)+A_2'\frac{a^3}{r^3}P_2(\cos\theta)+A_4'\frac{a^5}{r^5}P_4(\cos\theta)+,\text{ etc.}\right\}. \quad (91)$$

The coefficients must now be chosen to make this identical with the continuous flow solution (78), (79). The values so determined are

$$A_0'=\frac{2}{\pi}:A_2'=-\frac{1}{3}.\frac{2}{\pi}:A_4'=\frac{1}{5}\frac{2}{\pi}:\text{ and so on.} \quad . \quad . \quad . \quad (92)$$

Equation (91), with these values inserted, gives the periodic flow solution required for region $r>a$. To determine $F(\rho)$, we put $z=0$, $r=\rho$ in (91), (92), and obtain

$$v=\theta_0 e^{ik't}.F(\rho),$$

where
$$F(\rho).=\frac{2}{\pi}\left(\frac{a}{\rho}\right)^{\frac{1}{2}}\left\{\frac{K_{\frac{1}{2}}(i\lambda\rho)}{K_{\frac{1}{2}}(i\lambda a)}+\frac{1}{6}.\frac{K_{5/2}(i\lambda\rho)}{K_{5/2}(i\lambda a)}+,\text{ etc.}\right\}. \quad . \quad . \quad (93)$$

At the edge of the disk, $(\rho=a)$,

$$F(a)=\frac{2}{\pi}\left\{1+\frac{1}{6}+\frac{3}{40}+\frac{5}{112}+,\text{ etc.}\right\}, \quad . \quad . \quad . \quad . \quad (94)$$
$$=1.$$

in agreement with a well-known series for $\pi/2$.

The formal solution of the periodic flow problem is now expressible in the form of the sum of two Fourier integrals similar to (82) and (83) above,

$$v=\theta_0 e^{ik't}\int_0^\infty dk\, e^{-\sqrt{(k^2-\lambda^2)}z}\, k\mathrm{J}_0(k\rho)\int_0^a d\rho'\rho'\mathrm{J}_0(k\rho'), \quad . \quad . \quad . \quad (95)$$

and

$$v=\theta_0 e^{ik't}\int_0^\infty dk\, e^{-\sqrt{(k^2-\lambda^2)}z}\, k\mathrm{J}_0(k\rho)\int_a^\infty d\rho'\rho'\mathrm{J}_0(k\rho')\mathrm{F}(\rho'). \quad . \quad (96)$$

By performing a partial integration with respect to ρ' in the latter, and the complete integration with respect to ρ' in the former—the sum yields a single integral. But the study of the solution itself must meantime be left to a later opportunity.

The above complete solution of the problem is in agreement with an approximate solution for the case—a and λ very small—given by Lord Rayleigh ((6), p. 934).

[*General.*—In any case of periodic flow of fluid or of heat, it is obvious that if we make the frequency of the motion smaller and smaller the motion itself must approximate more and more closely to that corresponding to steady flow. This principle is employed by Rayleigh to obtain approximate solutions in several problems referred to above. It should be applicable to each individual term representing a periodic flow solution. According to this principle the relation between any periodic flow solution and the corresponding steady flow solutions is that represented by

$$v_c = \underset{\lambda=0}{\mathrm{Lt}}\, v_p. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (97)$$

This relation holds for all the corresponding solutions met with throughout the present investigation. Difficulties arise in the two cases mentioned in Part I. in which an infinite term occurs, namely in equations (10) and (17′) respectively. Thus, in the case of the line source solution (10),

$$\begin{aligned} v_c &= \underset{k=0}{\mathrm{Lt}}\, \frac{q}{2\pi\mathrm{K}} e^{ikt}\mathrm{K}_0(i\lambda\rho) \\ &= \frac{q}{2\pi\mathrm{K}}\, \underset{\lambda=0}{\mathrm{Lt}} -\log(i\lambda\rho) \\ &= -\frac{q}{2\pi\mathrm{K}}\log\rho\text{—an infinite term,} \quad . \quad . \quad . \quad . \quad (98) \end{aligned}$$

and in the case of the plane source solution (17′),

$$\begin{aligned} v_c &= \underset{\lambda=0}{\mathrm{Lt}}\, \frac{q}{2\mathrm{K}}\frac{e^{ikt-i\lambda x}}{i\lambda} \\ &= \underset{\lambda=0}{\mathrm{Lt}}\, \frac{q}{2\mathrm{K}}\cdot\frac{e^{ikt}}{i\lambda}(\cos\lambda x - i\sin\lambda x) \end{aligned}$$

$$=\text{infinite term } -\frac{q}{2K}x, \qquad (99)$$

in accordance with (97). It is evident that even in these cases the process and theorem of Part I. and the above principle yield like results ; in applications to physical problems the term containing the imaginary is usually of no significance. A mathematical proof—that the process and the principle are equivalent—is lacking. Of course, they are equivalent in the case of the point source solutions, and therefore also for multiple sources : but this does not explain clearly in what way the equivalence depends upon Fourier's integral Theorem.]

Narrow Rectangular Slit in a Thin Rigid Wall.—As in the case of the circular aperture of radius a, so in the case of a long slit of width $2a$, we have the corresponding problems of steady flow and of periodic flow through the aperture : and the solution of the periodic flow problem can be determined from that of the steady flow by applying exactly the same method as we have used in the problems relating to a circular aperture.

Continuous Flow Problem.—Taking the origin at the centre line of the slit, OZ perpendicular to the plane of the wall, the OX perpendicular to the edge of the slit and in the plane of the wall, the continuous flow problem is to determine the velocity potential function $\phi(z, x)$ to satisfy the conditions

$$\nabla^2\phi=0, \qquad (100)$$

$$\phi=\text{constant over slit } (-a<x<a), \qquad (101)$$

$$\frac{d\phi}{dz}=0, \text{ over the wall, } x>a,\ x<-a. \qquad (102)$$

In the analogous heat conduction problem, ϕ is to be replaced by v, the temperature at any point in the conducting medium surrounding an infinitely long thin plate of width $2a$, which is kept at a constant temperature θ_0. The first step in the procedure now being followed is to determine the steady flow solution to the heat conduction problem. This solution is readily obtained from a solution given by Lamb ('Hydrodynamics,' 3rd ed., § 66, 1°) of the problem of flow through an aperture of width $2a$, expressed in terms of conjugate functions in the form

$$x+iz=a\cosh w\,;\quad w=\phi+i\psi. \qquad (103)$$

By replacing w in this by $\frac{w-\theta_0}{\theta_0}\log\frac{a}{2}$, and then identifying ϕ so obtained with v, we obtain an expression for v which satisfies the conditions to be fulfilled in the present problem. This solution is exactly equivalent to that determined by Laplace's equation and the two conditions referring to the temperature distribution on the $z=0$ plane

$$v=\theta_0\ ;\ \text{on the plate } (-a<x<+a), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (104)$$

$$v=\theta_0\,.\,\mathrm{F}(x)\ ;\ \text{beyond the plate } (x<-a : x>a), \quad . \quad . \qquad (105)$$

where $\mathrm{F}(x)$ is defined by $(x>a)$.

$$v(x,\,0)=\frac{\theta_0}{\log\frac{a}{2}}\log\left(\frac{x+\sqrt{x^2-a^2}}{2}\right), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (106)$$

$$=\frac{\theta_0}{\log\frac{a}{2}}\left\{\log x-\left(\frac{1}{2}\,\frac{1}{2}\,\frac{a^2}{x^2}+\frac{1}{4}\,\frac{1\,.\,3}{2\,.\,4}\,\frac{a^4}{x^4}+\frac{1}{6}\,\frac{1\,.\,3\,.\,5}{2\,.\,4\,.\,6}\,\frac{a^6}{x^6}+,\ \text{etc.}\right)\right\}. \qquad (106')$$

The temperature distribution expressed by this equation is that required in the plane $z=0$, in the region external to the plate, in order that the normal gradient $\frac{dv}{dz}=0$ at plane $z=0$ for the complete solution. The expression (106) may also be obtained by an entirely different method.

Periodic Flow Problem.—In the corresponding periodic flow problem we have to find a solution of

$$(\nabla^2+\lambda^2)v=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (107)$$

$$v=\theta_0 e^{ikt},\ \text{on the plate}, \quad . \quad . \quad . \quad . \quad . \qquad (108)$$

and

$$\frac{dv}{dz}=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (109)$$

on plane $z=0$, external to the plate. This solution is determined by finding $\theta_0 e^{ikt}\mathrm{F}(x)$, the temperature distribution on plane $z=0$, in the region external to the plate of width $2a$, which must be imposed in order to fulfil condition (109).

As a general solution of (107) above, subject to the condition (109), we may write

$$v=\frac{\theta_0 e^{ikt}}{\log\frac{a}{2}}\left\{\sum_0^\infty \mathrm{A}_{2n}\frac{\mathrm{K}_{2n}(i\lambda\rho)}{\mathrm{K}_{2n}(i\lambda a)}\cos 2n\theta\right\}, \quad . \quad . \quad . \quad . \qquad (110)$$

where $\cos\theta=\frac{z}{\rho}$, and A_0, A_2, etc., are undetermined coefficients. Each term of this summation is itself a periodic flow solution of the wave-equation, and the corresponding steady flow solution may be derived from it by the general process employed above. By making use of relations stated in (89) and (90) above, we obtain $(\rho>a)$,

$$v=\frac{\theta_0}{\log\frac{a}{2}}\left\{\mathrm{A}_0\frac{\log\rho}{\log a}+\mathrm{A}_2\,.\,\frac{a^2}{\rho^2}\cos 2\theta+\mathrm{A}_4\,.\,\frac{a^4}{\rho^4}\cos 4\theta+,\ \text{etc.}\right\}, \quad . \quad . \qquad (111)$$

as steady flow solution corresponding to (110). The coefficients, A_0, A_4, etc., are now to be determined so that the steady flow solution (111) is identical with that obtained by means of Lamb's solution (103). Putting $z=0$, and $\theta=\frac{\pi}{2}$, in (111) we obtain $v(x, 0)$; and by identifying this with (106′) above we find

$$A_0=\log a : A_2=\frac{1}{2}\frac{1}{2} : A_4=-\frac{1}{4}\frac{1.3}{2.4} : A_6=\frac{1}{6}\cdot\frac{1.3.5}{2.4.6} : \quad . \quad (112)$$

and so on. These values are now to be inserted in (110), which then represents the periodic flow solution for the region $(\rho>a)$: and the temperature distribution on the $z=0$ plane $(x>a)$ is determined by

$$\begin{aligned} v(x, 0) &= \theta_0 e^{ikt} F(x) \\ &= \frac{\theta_0 e^{ikt}}{\log\frac{a}{2}}\left\{\log a \cdot \frac{K_0(i\lambda x)}{K_0(i\lambda a)} - \Sigma\right\}, \quad . \quad . \quad . \quad . \quad (113) \end{aligned}$$

where
$$\Sigma=\frac{1}{2}\cdot\frac{1}{2}\frac{K_2(i\lambda x)}{K_2(i\lambda a)}+\frac{1}{4}\cdot\frac{1.3}{2.4}\frac{K_4(i\lambda x)}{K_4(i\lambda a)}+\text{, etc.}$$

At the edge of the plate, $x=a$,

$$\begin{aligned} F(a) &= \left\{\log a-\left(\frac{1}{2}\frac{1}{2}+\frac{1}{4}\cdot\frac{1.3}{2.4}+\frac{1}{6}\cdot\frac{1.3.5}{2.4.6}+\text{, etc.}\right)\right\}\frac{1}{\log\frac{a}{2}} \\ &= \left\{\log a-\log 2\right\}\frac{1}{\log\frac{a}{2}}=1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (114) \end{aligned}$$

and $v(a, 0)=\theta_0 e^{ikt}$ as required.

Just as in the case of periodic flow through a circular aperture, the formal solution of the problem of periodic flow through the slit may now be expressed in the form of a Fourier integral. For the problem in heat conduction treated above, the solution

$$v=-\frac{\theta_0 e^{ikt}}{\pi}\int_{-a}^{+a} dx_1 \frac{d}{dz} K_0(i\lambda\rho'), \quad . \quad . \quad . \quad . \quad . \quad (115)$$

where
$$\rho'^2=\{z^2+(x-x_1)^2\}$$

imposes the temperature distribution $v=\theta_0 e^{ikt}$ upon the plate $(-a<x_1<a)$: and the solution

$$v=-\frac{\theta_0 e^{ikt}}{\pi}\int_{a}^{\infty} dx_1 \frac{d}{dz}\{K_0(i\lambda\rho')+K_0(i\lambda\rho'')\}F(x_1), \quad . \quad . \quad (116)$$

where
$$\rho''^2=\{z^2+(x+x_1)^2\}$$

imposes the temperature distribution $v=\theta_0 e^{ikt}\mathrm{F}(x)$ on the regions, $x>a$ and $x<-a$, of the plane $z=0$ beyond the plate. The sum of these two solutions gives the complete solution of the problem.

These solutions illustrate the application of the method of corresponding solutions to determine the periodic flow solution from the continuous flow solution. Problems of periodic motion of fluid due to periodic motion of a circular disc or narrow blade may be solved by the same method. The study in detail of the solutions obtained above, and of the analogous problems to which they refer, may be possible at a later time.

List of References.

(1) Phil. Mag. iii. pp. 784–800 (April 1927, Suppl.).
(2) *Ibid.* v. pp. 701-720 (April 1928).
(3) *Ibid.* xii. pp. 233–255 (August 1931, Suppl.).
(4) *Ibid.* ser. 7, xviii. pp. 625–640 (Oct. 1934).
(5) *Ibid.* ser. 7, xxi. pp. 922–934 (May 1936).
(6) *Ibid.* ser. 7, xxi. pp. 934–947 (May 1936).
(7) *Ibid.* ser. 7, xxii. pp. 1079–1088 (Dec. 1936).
(8) *Ibid.* ser. 7, xxix. pp. 100–102 (Jan. 1940).

XXII. *The Transient Behaviour of Four-Terminal Networks.*

By Louis A. Pipes, Ph.D., Graduate School of Engineering,
Harvard University, Cambridge, Massachusetts *.

[Received June 13, 1939.]

Abstract.

Starting with the canonical differential equations of a general, linear, bilateral, lumped network, and by the application of the Laplacian transformation and matrix algebra, the general theory of the transient behaviour of four-terminal networks is deduced. By the use of simple matrix multiplication, the general constants of complicated structures are determined, and the exposition is concluded with a discussion of the general smooth transmission line and the general ladder structure. A brief discussion of the Laplacian transformation is given in an appendix. In order to apply the theory to special cases, two tables are included. The first table gives the basic theorems for the manipulation of Laplacian transforms and the second table is a list of transforms of great use in the discussion of special cases. Many of the transforms given are scattered widely over the mathematical literature, and a compilation of these in tabular form should greatly facilitate the solution of certain transient problems.

I. *Introduction.*

There has been considerable work recently on the application of tensors, dyadics, and matrices to electric circuit theory by several investigators (1), (2), (3), (4). Another group of investigators has been clarifying the ideas of Heaviside's Operational Calculus, and placing it on the firm foundation of the Laplacian Transformation theory (5), (6), (7), (8), (9), (10), (12). It is the purpose of this discussion to bring both of these powerful methods to bear on the solution of transient problems in four-terminal networks of various types, including the smooth transmission line and the general ladder structure. When these methods are used, it will be seen that a certain simplicity results that proves most useful in analysing complex structures. A knowledge of the Laplacian Transformation such as presented in (12), and a knowledge of matrix algebra as presented in (4), will be assumed in the discussion.

---

* Communicated by the Author.

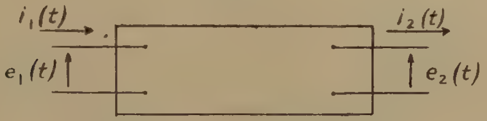
II. *The General Equations.*

Let us consider a general four-terminal network as shown in fig. 1. We will assume the reference directions for voltage and current as shown in the figure. The four-terminal network will be supposed to consist of an n mesh network, having only two accessible elements as shown. The impressed voltages $e_1(t)$ and $e_2(t)$ are arbitrary functions of time produced by sources external to the network, and the currents $i_1(t)$ and $i_2(t)$ are the mesh currents in meshes 1 and 2 produced by the application of the external electromotive forces inserted in the sense shown, in series with the two accessible terminals.

The internal structure of the four-terminal network will be assumed to be quite general, and the n meshes will be assumed to have mesh impedance operators of the form

$$Z_{mm}(D)=L_{mm}D+R_{mm}+S_{mm}\int_{-\infty}^{t}(\ )\,dt, \qquad D=\frac{d}{dt}, \quad . \quad . \quad (1)$$

Fig. 1.



General four-terminal network.

where L_{mm}, R_{mm}, and S_{mm} are the mesh inductance, resistance, and elastance coefficients respectively. The structure will also be assumed to have mutual impedance operators between meshes of the form

$$Z_{mn}(D)=L_{mn}D+R_{mn}+S_{mn}\int_{-\infty}^{t}(\ )\,dt, \quad . \quad . \quad . \quad . \quad (2)$$

where L_{mn}, R_{mn}, and S_{mn} are the coefficients of mutual inductance, resistance, and elastance respectively. The restrictions on the component elements of the network are that they be linear, constant and bilateral. It will be further assumed that the network is passive, that is, the only sources of electromotive force are $e_1(t)$ and $e_2(t)$ produced by the external agencies. The canonical differential equations of the network may be written in the following matrix form:

$$[Z(D)](i)=(e), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where (e) is a columnar matrix whose n elements are the n mesh voltages. $[Z(D)]$ is a square symmetric matrix of order n whose elements are the self and mutual operational impedances of n mesh network, and (i) is a columnar matrix of order n whose elements are the mesh currents of the network. In view of the symmetry conditions of the mutual coefficients of the system we have

$$Z_{mn}(D)=Z_{mn}(D), \quad \ldots \quad (4)$$

and, therefore,

$$[Z(D)]'=[Z(D)], \quad \ldots \quad (5)$$

where $[Z(D)]'$ is the transpose of the matrix $[Z(D)]$, and is obtained from $[Z(D)]$ by interchanging the rows and columns.

In the case under consideration, the system has accessible terminals in only two meshes, the input mesh which we will call mesh ♯1, and the output mesh which we will call mesh ♯2. Since the system is supposed devoid of internal electromotive forces, the voltage matrix has all its elements equal to zero with the exception of its first two.

III. *The Laplacian Transformation.*

The modern development of the Heaviside Calculus is based on the Fourier-Mellin theorem, which can be formulated in the following manner :

Given the real function $h(t)$ of the real variable t. which possesses the following properties :—

1. $h(t)$ and its first derivatives are piece-wise, continuous in each finite interval.
2. $h(t)=[h(t+0)+h(t-0)]/2$ at a point of discontinuity.
3. $\int_{-\infty}^{\infty}|h(t)|e^{-ct}dt$ exists for some real positive c.

Then $h(t)$ may be represented by the complex integral

$$h(t)=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\frac{e^{pt}g(p)\,dp}{p} \qquad j=\sqrt{-1}, \quad \ldots \quad (6)$$

where
$$g(p)=p\int_{-\infty}^{+\infty}h(t)e^{-pt}dt \quad \ldots \quad (7)$$

and
$$\text{real part of } p \geqslant c. \quad \ldots \quad (8)$$

If $h(t)=0$ for $t<0$, then we may place the lower limit of (7) equal to zero. We may then write

$$g(p)=p\int_{0}^{\infty}h(t)e^{-pt}dt. \quad \ldots \quad (9)$$

A detailed account of this theorem and its applications is given in (12). Some of the more important relations based on the integrals (6) and (9) will be found in the Appendix and Table I.

In the subsequent work, two functions $h(t)$ and $g(p)$ related by the equations (6) and (9) will be called the transforms of each other, and we shall use van der Pol's notation

$$h(t) \doteqdot g(p) \quad \ldots \ldots \ldots \quad (10)$$

to signify the functional relation between the functions, see ⁽⁷⁾.

To introduce the Laplacian Transformation into the problem under consideration, let us multiply the matrix equation (3) by $pe^{-pt}dt$, and integrate both sides from 0 to infinity. We thus obtain

$$p\int_0^\infty e^{-pt}[\mathrm{Z(D)}](i)\,dt = p\int_0^\infty e^{-pt}(e)\,dt. \quad \ldots \ldots \quad (11)$$

Now, if we let

$$(\mathrm{E}) = p\int_0^\infty e^{-pt}(e)\,dt \quad \ldots \ldots \ldots \quad (12)$$

and

$$(\mathrm{I}) = p\int_0^\infty e^{-pt}(i)\,dt, \quad \ldots \ldots \ldots \quad (13)$$

and integrate (11) by parts, we obtain

$$[\mathrm{Z}(p)](\mathrm{I}) = (\mathrm{E}) = p[\mathrm{L}]i^0) - [\mathrm{S}](q^0), \quad \ldots \ldots \quad (14)$$

where the elements of the columnar matrix (I) are the transforms of the elements of the matrix (i) as given by (13). (i^0) is a columnar matrix whose elements are the initial mesh currents of the network. (q^0) is a columnar matrix whose elements are the initial mesh charges of the system. The square matrix $[\mathrm{Z}(p)]$ is given by

$$[\mathrm{Z}(p)] = p[\mathrm{L}] + [\mathrm{R}] + [\mathrm{S}]/p, \quad \ldots \ldots \ldots \quad (15)$$

where [L], [R], and [S] are square matrices whose elements are the inductance, resistance, and elastance coefficients respectively.

We shall assume that before the application of the arbitrary electromotive forces at $t=0$, the system was devoid of mesh charges and mesh currents. In such a case we must have

$$(i^0) = (0) \quad \ldots \ldots \ldots \quad (16)$$

and

$$(q^0) = (0). \quad \ldots \ldots \ldots \quad (17)$$

The transformed equation (13) then reduces to

$$[\mathrm{Z}(p)](\mathrm{I}) = (\mathrm{E}). \quad \ldots \ldots \ldots \quad (18)$$

Since the system is supposed devoid of internal electromotive forces, the transformed voltage matrix (E) has all its elements equal to zero with the exception of the first two. If we pre-multiply (18) by $[\mathrm{Z}(p)]^{-1}$, the inverse of $[\mathrm{Z}(p)]$, we have an explicit relation for the transform of the current matrix

$$(\mathrm{I}) = [\mathrm{Z}(p)]^{-1}(\mathrm{E}). \quad \ldots \ldots \ldots \quad (19)$$

Let us introduce a square matrix of the nth order $[Y(p)]$ defined by the equation

$$[Y(p)]=[Z(p)]^{-1}. \quad \ldots \quad (20)$$

Since $[Y(p)]$ is the inverse of $[Z(p)]$ and $[Z(p)]$ is a symmetric matrix, we have :

$$Y_{mn}(p)=\frac{\text{Cofactor of } Z_{mn}(p)}{Z(p)}, \quad \ldots \quad (21)$$

where $Y_{mn}(p)$ is the general element of the matrix $[Y(p)]$ and $Z(p)$ is the determinant of the matrix $[Z(p)]$. Hence we may write concisely

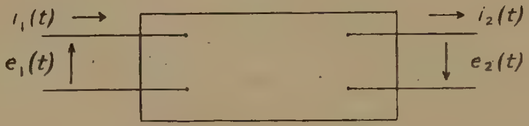
$$(I)=[Y(p)](E). \quad \ldots \quad (22)$$

From (22), since we are only interested in the two accessible currents $i_1(t)$ and $i_2(t)$, we write

$$\left.\begin{aligned} I_1(p)&=Y_{11}(p)E_1(p)+Y_{12}(p)E_2(p),\\ I_2(p)&=Y_{21}(p)E_1(p)+Y_{22}(p)E_2(p),\end{aligned}\right\} \quad \ldots \quad (23)$$

for their transforms.

Fig. 2.



In the above, it has been assumed that positive currents in each mesh have clockwise directions and that positive electromotive forces tend to circulate positive currents in the meshes where they are applied. It is easy to see that with this convention of signs, all the mutual operators $Z_{mn}(p)$ are negative. Now, in the theory of four-terminal networks, it is convenient to use a somewhat different reference direction for the voltage $e_2(t)$. The standard conventional directions are shown in fig. 2.

The only alteration necessary in the above equations to make them conform to the standard reference direction, as shown in fig. 2, is the reversal of the sign of $e_2(t)$, and hence that of its transform $E_2(p)$. Hence, if we write

$$\left.\begin{aligned} I_1(p)&=Y_{11}(p)E_1(p)-Y_{12}(p)E_2(p),\\ I_2(p)&=Y_{21}(p)E_1(p)-Y_{22}(p)E_2(p),\end{aligned}\right\} \quad \ldots \quad (24)$$

we conform to the standard reference directions of fig. 2.

Solving for $E_1(p)$ and $I_1(p)$ in equations (24) we arrive at the following form of the equations :

$$\left.\begin{aligned} E_1(p)&=A(p)E_2(p)+B(p)I_2(p),\\ I_1(p)&=C(p)E_2(p)+D(p)I_2(p),\end{aligned}\right\} \quad \ldots \quad (25)$$

where

$$\left.\begin{aligned} A(p)&=Y_{22}/Y_{21}, & B(p)&=1/Y_{11}, \\ C(p)&=(Y_{11}Y_{22}-Y_{12}^2)/Y_{12}, & D(p)&=Y_{11}/Y_{12}. \end{aligned}\right\} \quad . \quad . \quad . \quad (26)$$

By a direct multiplication, it is seen that we have the identity

$$AD-BC=1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

The equations (25) may be written in the convenient matrix form

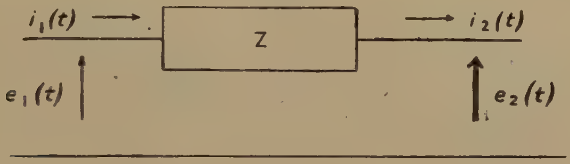
$$\begin{bmatrix} E_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} E_2 \\ I_2 \end{bmatrix}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

The general theory thus shows that any linear, passive, bilateral, four-terminal network is characterized by a square matrix of the second order whose determinant is unity. This matrix,

$$[u] = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

will be called the " associated matrix " of the network.

Fig. 3.



IV. *The Associated Matrices of Common Four-Terminal Structures.*

We shall now consider how the associated matrices of complicated structures may be obtained from a consideration of combinations of simple structures and matrix multiplication.

A. *Impedance in Series.*

If we have an impedance in series, as shown in fig. 3, we have by inspection its associated matrix given by (30),

$$[u] = \begin{bmatrix} 1 & Z(p) \\ 0 & 1 \end{bmatrix}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

B. *Shunt Admittance.*

Similarly, (31) is the associated matrix of the shunt admittance of fig. 4,

$$[u] = \begin{bmatrix} 1 & 0 \\ Y(p) & 1 \end{bmatrix}.$$

C. *T-Network.*

The T-network of fig. 5 may be regarded as the cascade connexion of a series impedance Z_a, a shunt admittance Y_c, and a series impedance Z_b. The associated matrix of the entire circuit is obtained by the matrix multiplication (32),

$$[u]=\begin{bmatrix}1 & Z_a\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ Y_c & 1\end{bmatrix}\begin{bmatrix}1 & Z_b\\ 0 & 1\end{bmatrix}=\begin{bmatrix}(1+Z_aY_c) & (Z_b+Z_aZ_bY_c+Z_a)\\ Y_c & (Z_bY_c+1)\end{bmatrix}. \quad (32)$$

Fig. 4.

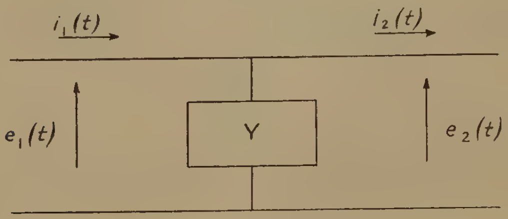


Fig. 5.

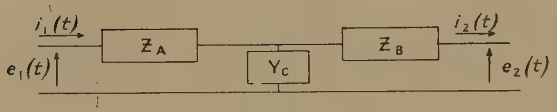
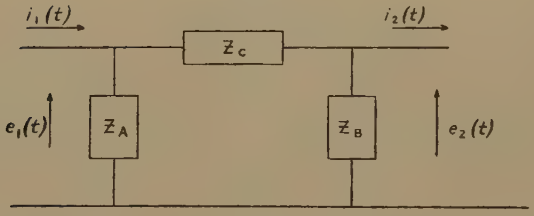


Fig. 6.



D. *Pi-Network.*

The Pi-network of fig. 6 may be regarded as the cascade connexion of a shunt impedance Z_a, a series impedance Z_c, and a shunt impedance Z_b. Its associated matrix $[u]$ may therefore be obtained by the following matrix multiplication of (33):

$$\left.\begin{aligned}[u]&=\begin{bmatrix}1 & 0\\ 1/Z_a & 1\end{bmatrix}\begin{bmatrix}1 & Z_c\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 1/Z_b & 1\end{bmatrix}\\ &=\begin{bmatrix}(Z_b+Z_c)/Z_b & Z_c\\ (Z_a+Z_b+Z_c)/Z_aZ_b & (Z_a+Z_c)/Z_a\end{bmatrix}\end{aligned}\right\} \quad . \; . \; . \; . \quad (33)$$

Similarly, the associated matrices of other structures may be obtained by direct matrix multiplication. We shall have occasion to do this in a later section, when we come to consider the transient behaviour of filter circuits and the smooth transmission line.

V. *Four-Terminal Network Terminated by an Impedance.*

Let us consider the problem of determining the current and voltage distribution of a four-terminal network terminated by a general impedance, as shown in fig. 7.

The system has impressed upon it an arbitrary electromotive force $e_1(t)$, and the problem is to determine the unknowns $i_1(t)$, $i_2(t)$, and $e_2(t)$ when the associated matrix of the four-terminal network $[u]$ and the impedance Z are known. By our general theory, we have

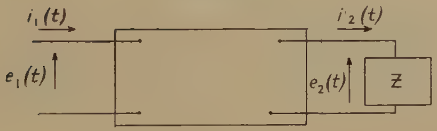
$$e_1(t) \doteqdot \mathrm{E}_1(p), \qquad (34)$$

$$e_2(t) \doteqdot \mathrm{E}_2(p), \qquad (35)$$

$$i_1(t) \doteqdot \mathrm{I}_1(p), \qquad (36)$$

$$i_2(t) \doteqdot \mathrm{I}_2(p). \qquad (37)$$

Fig. 7.



If the system has no initial charges or initial currents before $t=0$, the transforms satisfy the equation (28),

$$\begin{bmatrix} \mathrm{E}_1 \\ \mathrm{I}_1 \end{bmatrix} = \begin{bmatrix} \mathrm{A}(p) & \mathrm{B}(p) \\ \mathrm{C}(p) & \mathrm{D}(p) \end{bmatrix} \begin{bmatrix} \mathrm{E}_2 \\ \mathrm{I}_2 \end{bmatrix} \qquad (28)$$

and

$$\mathrm{E}_2 = \mathrm{Z}(p)\mathrm{I}_2. \qquad (38)$$

Hence, if we substitute this value of E_2 in (28), we obtain

$$\begin{bmatrix} \mathrm{E}_1 \\ \mathrm{I}_1 \end{bmatrix} = \begin{bmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{C} & \mathrm{D} \end{bmatrix} \left| \begin{matrix} \mathrm{Z} \\ 1 \end{matrix} \right| \mathrm{I}_2. \qquad (39)$$

If we call $\mathrm{Z}_1(p)$ the operational impedance, looking into the network from the left when it is terminated as shown, we have

$$\mathrm{Z}_1(p) = \mathrm{E}_1/\mathrm{I}_1 = (\mathrm{AZ}+\mathrm{B})/(\mathrm{CZ}+\mathrm{D}). \qquad (40)$$

Since $\mathrm{A}(p)$, $\mathrm{B}(p)$, $\mathrm{C}(p)$, $\mathrm{D}(p)$, and $\mathrm{Z}(p)$ are supposed known, we compute $\mathrm{Z}_1(p)$ from (40) ; we then have

$$\mathrm{I}_1(p) = \mathrm{E}_1(p)/\mathrm{Z}_1(p). \qquad (41)$$

Since the electromotive force $e_1(t)$ is supposed known, its transform may be computed from (9) or from the table of transforms, hence $E_1(p)$ is known. The response current $i_1(t)$ is thus given by the basic integral (6) in the form

$$i_1(t)=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\frac{e^{pt}E_1(p)\,dp}{pZ_1(p)}. \qquad (42)$$

This integral may be evaluated by the theory of residues when the poles of the integrand are known. In many cases, however, the inverse transform of $I_1(p)$ may be read directly from Table II.

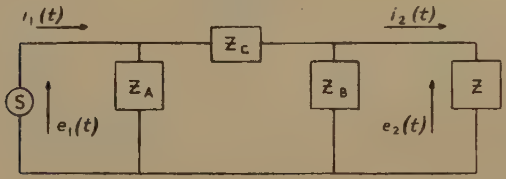
To determine $E_2(p)$ and $I_2(p)$, pre-multiply (28) by $[u]^{-1}$, we then have

$$\begin{bmatrix} E_2 \\ I_2 \end{bmatrix}=\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} E_1 \\ I_1 \end{bmatrix}. \qquad (43)$$

In view of the equation (27) we have

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1}=\begin{bmatrix} D & -B \\ -C & A \end{bmatrix}. \qquad (44)$$

Fig. 8.



Therefore, we have

$$\begin{bmatrix} E_2 \\ I_2 \end{bmatrix}=\begin{bmatrix} D & -B \\ -C & A \end{bmatrix}\begin{bmatrix} E_1 \\ I_1 \end{bmatrix}. \qquad (45)$$

Substituting (41) into (45) we have

$$\begin{bmatrix} E_2 \\ I_2 \end{bmatrix}=\begin{bmatrix} D & -B \\ -C & A \end{bmatrix}\begin{bmatrix} 1 \\ 1/Z_1 \end{bmatrix}E_1, \qquad (46)$$

or multiplying, we obtain

$$E_2=DE_1-BE_1/Z_1, \qquad (47)$$

$$I_2=AE_1/Z_1-CE_1, \qquad (48)$$

and from the fundamental integral, we obtain

$$i_2(t)=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\frac{e^{pt}I_2(p)\,dp}{p} \qquad (49)$$

and

$$e_2(t)=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\frac{e^{pt}E_2(p)\,dp}{p}. \qquad (50)$$

As an illustrative example, let us consider the case of the Pi-network of fig. 6, terminated by an impedance Z as shown in fig. 8.

From (33) we have

$$A(p)=(Z_b+Z_c)/Z_b, \quad \ldots \quad (51)$$

$$B(p)=Z_c, \quad \ldots \quad (52)$$

$$C(p)=(Z_a+Z_b+Z_c)/Z_aZ_b, \quad \ldots \quad (53)$$

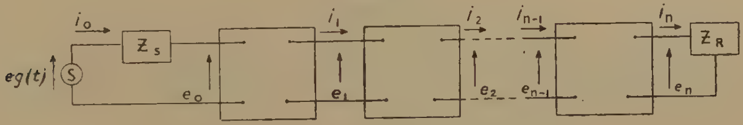
$$D(p)=(Z_a+Z_c)/Z_a. \quad \ldots \quad (54)$$

Hence $Z_1(p)$ is given by (40) and the response currents and voltages by (42), (49), and (50). The ease of evaluation of these integrals depends upon the nature of the poles of the integrands. In the general case, the zeros of $Z_1(p)$ are simple zeros, and their number is in general equal to twice the number of closed meshes of the entire system. However, special cases arise where this rule does not hold, depending on the character of the component impedances of the system.

VI. *Cascade Connexion of Identical Four-Terminal Networks.*

Let us consider the problem of determining the voltages and currents of a system composed of n identical dissymetrical four-terminal networks arranged as shown in fig. 9.

Fig. 9.



This problem includes the general filter circuit problem in the transient state. Throughout the analysis we shall use the notation

$$i_r(t) \doteqdot I_r(p) \quad (r=0, 1, \ldots n), \quad \ldots \quad (55)$$

$$e_r(t) \doteqdot E_r(p), \quad \ldots \quad (56)$$

$$e_g(t) \doteqdot E_g(p). \quad \ldots \quad (57)$$

Since the four-terminal networks are identical, we may regard the arrangement as a chain of n identical circuits connected in cascade with a series impedance Z_s. Accordingly, we have for the transforms the general relation

$$\begin{bmatrix} E_g \\ I_0 \end{bmatrix} = \begin{bmatrix} 1 & Z_s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix}^n \begin{bmatrix} E_n \\ I_n \end{bmatrix}. \quad \ldots \quad (58)$$

In this equation, we could, of course, raise the matrix $[u]$ to any integral power n by direct multiplication. Such a procedure is tedious, however, and it is better to make use of the basic theorems of matrix algebra.

As a first step in the procedure of raising the matrix $[u]$ to the integral power n, let us construct the characteristic matrix of the matrix $[u]$ defined by the equation

$$[f(m)]=\begin{bmatrix}(m-\mathrm{A}) & -\mathrm{B} \\ -\mathrm{C} & (m-\mathrm{D})\end{bmatrix}. \qquad (59)$$

The characteristic equation of $[u]$ is in this case

$$f(m)\,|=m^2-m(\mathrm{A}+\mathrm{D})+1=0, \qquad (60)$$

and the latent roots of $[u]$ are

$$m_{1,\,2}=(\mathrm{A}+\mathrm{D})/2\pm\sqrt{(\mathrm{A}+\mathrm{D})^2/4-1}. \qquad (61)$$

We note that the product of the latent roots satisfies the equation

$$m_1 . m_2=1, \qquad (62)$$

and the sum of the latent roots satisfies the equation

$$m_1+m_2=(\mathrm{A}+\mathrm{D}). \qquad (63)$$

Accordingly, if we let

$$m_1=e^a, \qquad (64)$$

where a is a function of p, we have, in view of (62),

$$m_2=e^{-a}. \qquad (65)$$

Now, as a consequence of the relation (63), we have

$$\cosh\,(a)=(\mathrm{A}+\mathrm{D})/2. \qquad (66)$$

The adjoint matrix of the characteristic matrix (59) is

$$[\mathrm{F}(m)]=\begin{bmatrix}(m-\mathrm{D}) & \mathrm{B} \\ \mathrm{C} & (m-\mathrm{A})\end{bmatrix}. \qquad (67)$$

Let us form the vector $(k)_1$ from the matrix (67) by substituting $m=m_1$ into its first column, that is, we form

$$(k)_1=\begin{Bmatrix}(e^a-\mathrm{D}) \\ \mathrm{C}\end{Bmatrix}. \qquad (68)$$

Similarly, let us form the vector $(k)_2$ by substituting $m=m_2$ into the first column of (67), that is

$$(k)_2=\begin{Bmatrix}(e^{-a}-\mathrm{D}) \\ \mathrm{C}\end{Bmatrix}. \qquad (69)$$

We now form the square matrix $[k]$ from the two vectors $(k)_1$ and $(k)_2$, so that we have

$$[k]=\begin{bmatrix}(e^a-\mathrm{D}) & (e^{-a}-\mathrm{D}) \\ \mathrm{C} & \mathrm{C}\end{bmatrix}. \qquad (70)$$

Now, let

$$s_1=(e^a-\mathrm{D})/\mathrm{C}, \qquad (71)$$

$$s_2=(e^{-a}-\mathrm{D})/\mathrm{C}, \qquad (72)$$

we then have

$$[k]=C\begin{bmatrix} s_1 & s_2 \\ 1 & 1 \end{bmatrix}. \quad \ldots \ldots \quad (73)$$

Computing the inverse of the square matrix $[k]$, we have

$$[k]^{-1}=\frac{1}{C(s_1-s_2)}\begin{bmatrix} 1 & -s_2 \\ -1 & s_1 \end{bmatrix}. \quad \ldots \ldots \quad (74)$$

Now, there is a theorem in matrix algebra $^{(4)}$, which states that any matrix $[u]$ may, by post-multiplication and pre-multiplication by the matrices $[k]$ and $[k]^{-1}$ constructed as above, be written in the form

$$[u]=[k]\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}[k]^{-1}, \quad \ldots \ldots \quad (75)$$

provided that the two latent roots m_1 and m_2 are distinct. In this case, this imposes the restriction that

$$(A+D)\neq\pm 2. \quad \ldots \ldots \quad (76)$$

The canonical form (75) is well suited for raising the matrix $[u]$ to a power, since by direct multiplication it may be easily shown that

$$[u]^r=[k]\begin{bmatrix} m_1^r & 0 \\ 0 & m_2^r \end{bmatrix}[k]^{-1}, \quad \ldots \ldots \quad (77)$$

where r is a positive or negative integer. By a direct substitution of the relations (61) into (71) and (72), it may be easily shown that

$$s_1=\frac{(A-D)+\sqrt{(A+D)^2-4}}{2C}, \quad \ldots \ldots \quad (78)$$

$$s_2=\frac{(A-D)-\sqrt{(A+D)^2-4}}{2C}. \quad \ldots \ldots \quad (79)$$

In view of (78) and (79), the quantities s_1 and s_2 are seen to be related to the quantities Z_{01} and Z_{02} of the literature by the equations

$$s_1=Z_{01}, \quad \ldots \ldots \quad (80)$$

$$-s_2=Z_{02}, \quad \ldots \ldots \quad (81)$$

where $Z_{01}(p)$ is the operational iterative impedance of the four-terminal networks in the positive sense and Z_{02} is the operational iterative impedance of the chain of circuits in the negative sense. Substituting (80) and (81) into (77) and carrying out the matrix multiplication, we obtain

$$[u]^r=\frac{1}{(Z_{01}+Z_{02})}\begin{bmatrix} (Z_{01}e^{ar}+Z_{02}e^{-ar}) & (2Z_{01}Z_{02}\sinh ar) \\ 2\sinh(ar) & (Z_{02}e^{ar}+Z_{01}e^{-ar}) \end{bmatrix}, \quad . \quad (82)$$

where r is a positive or negative integer.

If the individual quadripoles of fig. 9 are symmetric, we have

$$A=D, \qquad \ldots \ldots \ldots \tag{83}$$

$$Z_{01}=Z_{02}. \qquad \ldots \ldots \ldots \tag{84}$$

In this case (82) reduces to :

$$[u]^r=\begin{bmatrix} \cosh (ar) & Z_{01} \sinh (ar) \\ \dfrac{\sinh (ar)}{Z_{01}} & \cosh (ar) \end{bmatrix}. \qquad \ldots \ldots \tag{85}$$

Having reduced the matrix $[u]^r$ to the convenient form (82), we now proceed to the investigation of the various current and voltage distributions of the system. For simplicity, let us introduce the notation

$$[u]^r=\begin{bmatrix} A_r & B_r \\ C_r & D_r \end{bmatrix}. \qquad \ldots \ldots \tag{86}$$

We then have from (58) the equation

$$\begin{bmatrix} E_g \\ I_0 \end{bmatrix}=\begin{bmatrix} 1 & Z_s \\ 0 & 1 \end{bmatrix}\begin{bmatrix} A_n & B_n \\ C_n & D_n \end{bmatrix}\begin{bmatrix} E_n \\ I_n \end{bmatrix}. \qquad \ldots \ldots \tag{87}$$

Now, since we have the relation

$$E_n=I_nZ_r, \qquad \ldots \ldots \ldots \tag{88}$$

by substituting this into (87) we obtain the following equation for the operational impedance of the entire system when looking at it from the generator end :

$$Z_g=E_g/I_0=\frac{(A_n+Z_sC_n)Z_r+(B_n+Z_sD_n)}{(C_nZ_r+D_n)}. \qquad \ldots \ldots \tag{89}$$

To find the transforms of the voltage and current at the rth unit along the cascade connexion of fig. 9, we write

$$\begin{bmatrix} E_g \\ I_0 \end{bmatrix}=\begin{bmatrix} 1 & Z_s \\ 0 & 1 \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} E_r \\ I_r \end{bmatrix}. \qquad \ldots \ldots \tag{90}$$

On finding the inverse of the resultant square matrix on the right side of (90), we have

$$\begin{bmatrix} E_r \\ I_r \end{bmatrix}=\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} 1 & -Z_s \\ 0 & 1 \end{bmatrix}\begin{bmatrix} E_g \\ I_0 \end{bmatrix}. \qquad \ldots \ldots \tag{91}$$

In view of (86) and (89) this may be written in the form

$$\begin{vmatrix} E_r \\ I_r \end{vmatrix}=\begin{bmatrix} A_{-r} & B_{-r} \\ C_{-r} & D_{-r} \end{bmatrix}\begin{bmatrix} 1 & -Z_s \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1/Z_g \end{bmatrix}E_g. \qquad \ldots \tag{92}$$

In (92) we have the transforms E_r and I_r expressed directly in terms of the transform $E_g(p)$ of the known electromotive force $e_g(t)$. Substituting the relation for Z_g as given in (89), and the elements of the matrix $[u]^{-r}$ as given by (82) into (92), we obtain, after some algebraic reductions, the following relations for the transforms of the voltages and currents along the line

$$E_k(p)=\frac{E_g(p)[Z_{01}e^{(n-k)a}+Z_{02}r_re^{-(n-k)a}]}{(Z_s+Z_{01})(e^{na}-r_sr_re^{-na})}, \qquad (93)$$

$$I_k(p)=\frac{E_g(p)[e^{(n-k)a}-r_re^{-(n-k)a}]}{(Z_s+Z_{01})(e^{na}-r_sr_re^{-na})}, \qquad (94)$$

where

$$r_s=(Z_s-Z_{02})/(Z_s+Z_{01}), \qquad (95)$$

$$r_r=(Z_r-Z_{01})/(Z_r+Z_{02}). \qquad (96)$$

This is the conventional form of the solution for the transforms $E_k(p)$ and $I_k(p)$. It may be noted that this solution was obtained by direct matrix multiplication in a most direct manner. The voltages and currents at any point along the system are then given by the equations

$$e_k(t)=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\frac{e^{pt}E_k(p)\,dp}{p}, \qquad (97)$$

$$i_k(t)=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\frac{e^{pt}I_k(p)\,dp}{p}. \qquad (98)$$

In the general case, the evaluation of the integrals (97) and (98) is a formidable task. The procedure is greatly facilitated by using the transforms of Table II. in specific cases. A procedure which enables full use of the transforms of Table II. to be made in quite general cases is the following one :—

We may write

$$\frac{1}{e^{na}-r_sr_re^{-na}}=(m_1^n-r_sr_rm_2^n)^{-1}=m_2^n(1-r_sr_rm_2^{2n})^{-1} \qquad (99)$$

in view of the relations (64) and (65).

This term may be written in the form

$$m_2^n\sum_{q=0}^{\infty}(r_sr_rm_2^{2n})^q. \qquad (100)$$

by a formal expansion by the binomial theorem. Multiplying this expression by the remaining terms of (93) we obtain

$$E_k(p)=\frac{E_gZ_{01}}{(Z_s+Z_{01})}[m_2^k+r_sr_rm_2^{(2n+k)}+(r_sr_r)^2m_2^{(4n+k)}$$
$$+(r_sr_r)^3m_2^{(6n+k)}+\ldots+]+\frac{E_gZ_{02}r_r}{(Z_s+Z_{01})}[m_2^{(2n-k)}+r_sr_rm_2^{(4n-k)}$$
$$+(r_sr_r)^2m_2^{(6n-k)}+(r_sr_r)^3m_2^{(8n-k)}+\ldots+]. \qquad (101)$$

The terms in the first group are the transforms of voltage waves travelling from the sending end toward the receiving end, while the terms in the second group are the transforms of waves travelling from the receiving end back to the sending end. It may be shown that in computing the inverse transforms of the terms of the series (101) by the basic integral (6) the series is absolutely convergent, and term by term integration is justified. Therefore, the expansion is legitimate. Repeating the procedure for the transform of the current (94), one obtains

$$\begin{aligned}\mathrm{I}_k(p)=&\frac{\mathrm{E}_g}{(\mathrm{Z}_s+\mathrm{Z}_{01})}[m_2^k+r_sr_rm_2^{(2n+k)}+(r_sr_r)^2m_2^{(4n+k)}\\&\qquad+(r_sr_r)^3m_2^{(6n+k)}+\ldots+]\\&-\frac{\mathrm{E}_gr_r}{(\mathrm{Z}_s+\mathrm{Z}_{01})}[m_2^{(2n-k)}+r_sr_rm_2^{(4n-k)}+(r_sr_r)^2m_2^{(6n-k)}\\&\qquad+(r_sr_r)^3m_2^{(8n-k)}+\ldots+].\end{aligned}\qquad(102)$$

The two groups of terms have the significance of being the transforms of current waves travelling to the right and to the left of the chain of circuits. To obtain the inverse transforms of (101) and (102) we set up the following sequence :

$$v_q(t)\doteqdot\frac{\mathrm{E}_g\mathrm{Z}_{01}(r_sr_r)^q}{(\mathrm{Z}_s+\mathrm{Z}_{01})}\quad(q=0,1,2,\ldots),\qquad(103)$$

$$a_q(t)\doteqdot pm_2^q,\qquad(104)$$

$$w_q(t)\doteqdot\frac{\mathrm{E}_g\mathrm{Z}_{02}r_r(r_sr_r)^q}{(\mathrm{Z}_s+\mathrm{Z}_{01})}\quad(q=0,1,2).\qquad(105)$$

Then, in view of the fundamental Faltung theorem of the operational calculus, which may be stated in the form :

If $$g_1(p)\doteqdot h_1(t)\quad\text{and}\quad g_2(p)\doteqdot h_2(t),\qquad(106)$$

then $$\begin{aligned}g_1(p)g_2(p)/p&\doteqdot\int_0^t h_1(u)h_2(t-u)\,du\\&\doteqdot\int_0^t h_2(u)h_1(t-u)\,du,\end{aligned}\qquad(107)$$

we have

$$\left.\begin{aligned}e_k(t)=&\sum_{q=0}^{\infty}\int_0^t v_q(t-u)a(u)_{(2qn+k)}\,du\\&+\sum_{q=0}^{\infty}\int_0^t w_q(t-u)a(u)_{[n(2+q)-k]}\,du.\end{aligned}\right\},\qquad(108)$$

Similarly, if we define the following sequence :

$$x_q(t)\doteqdot\frac{\mathrm{E}_g(r_sr_r)^q}{(\mathrm{Z}_s+\mathrm{Z}_{01})}\quad(q=0,1,2,\ldots),\qquad(109)$$

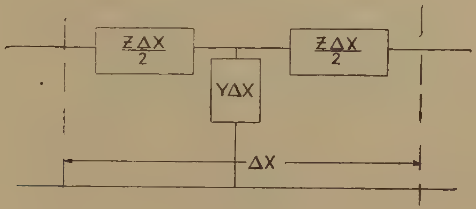
$$y_q(t)\doteqdot\frac{\mathrm{E}_gr_r(r_sr_r)^q}{(\mathrm{Z}_s+\mathrm{Z}_{01})}.\qquad(110)$$

then, we have

$$\left.\begin{aligned} i_k(t) = & \sum_{q=0}^{\infty} \int_0^t x_q(t-u) a(u)_{(2qn+k)} \, du \\ & - \sum_{q=0}^{\infty} \int_0^t y_q(t-u) a(u)_{[n(2+q)-k]} \, du. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (111)$$

The most feasible method of solution in special cases is to consider E_g to be a steady electromotive force of one unit and then to obtain the response to an arbitrary electromotive force by the principle of superposition or the Faltung theorem. Transforms that enable a great many special problems to be solved in closed form are given in Table II. In general, if the circuit has resistances, the reflected waves diminish greatly in magnitude after a few successive reflexions, hence it is only necessary to take a few terms of the series (108) and (111). The computation of the inverse transforms of the sequences (103), (104), (105), (109) and (110) are greatly facilitated by the use of the table of transforms in the Appendix. If the impedance of the terminal apparatus satisfies the equation

Fig. 10.



$$Z_r = Z_{01}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (112)$$

then

$$r_r = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (113)$$

and the series (108) and (111) both reduce to a single term. This case of matched impedances prevents successive reflexions from being established and results in a tremendous simplification of the problem.

VII. *The Smooth Transmission Line.*

We shall now regard the smooth transmission line as the limiting case of an infinite number of symmetrical four-terminal networks. Each of these symmetrical networks will be of the type shown in fig. 10.

Let us consider a line having an operational series impedance $Z(p)$ per unit length and an operational shunt admittance $Y(p)$ per unit length. Then we may regard the line as made up of an infinite number of four-terminal networks of the form given in fig. 10. The associated matrix of the network of fig. 10 is

$$[u]=\begin{bmatrix}(1+ZY\Delta x^2/2) & (Z\Delta x+Z^2Y\Delta x^3/4)\\ Y\Delta x & (1+ZY\Delta x^2/2)\end{bmatrix}. \quad . \quad . \quad . \qquad (114)$$

Since the infinitesimal network of fig. 10 is symmetric, we have

$$A=D. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (115)$$

In this case, by formulæ (78) and (79) we have

$$Z_{01}=Z_{02}=Z_0=\frac{\sqrt{(A+D)^2-4}}{2C} \quad . \quad . \quad . \quad . \quad . \quad . \qquad (116)$$

for the characteristic impedance of the line. Substituting the values of A, D, and C from (114) into this equation, and neglecting higher order terms, we obtain

$$Z_0=\sqrt{Z(p)/Y(p)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (117)$$

To compute the propagation constant, we use the equation (66),

$$\cosh(a)=(A+D)/2=1+ZY\Delta x^2/2.$$

Expanding cosh (a) in a Taylor's series, we have

$$1+a^2/2!+a^4/4!+\ldots=1+ZY\Delta x^2/2!. \quad . \quad . \quad . \qquad (118)$$

Hence we have

$$a=\Delta x\sqrt{ZY}=\Delta xb. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (119)$$

Substituting this into (82), we have

$$[u]^n=\begin{bmatrix}\cosh(n\Delta x\sqrt{ZY}) & (\sqrt{Z/Y})\sinh(n\Delta x\sqrt{ZY})\\ (\sqrt{Y/Z})\sinh(n\Delta x\sqrt{ZY}) & \cosh(n\Delta x\sqrt{ZY})\end{bmatrix}. \qquad (120)$$

On passing to the limit we make use of the relation

$$\underset{\substack{n\to\infty\\ \Delta x\to 0}}{\text{Limit}}\,(n\Delta x)=s, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (121)$$

where s is the length of the line. Now if we substitute these relations into (93) and (94), we have

$$E(x,p)=\frac{Z_0E_g[e^{b(s-x)}+r_re^{-b(s-x)}]}{(Z_s+Z_0)(e^{bs}-r_sr_re^{-bs})}, \quad . \quad . \quad . \quad . \quad . \qquad (122)$$

$$I(x,p)=\frac{E_g[e^{b(s-x)}-r_re^{-b(s-x)}]}{(Z_s+Z_0)(e^{bs}-r_sr_re^{-bs})}, \quad . \quad . \quad . \quad . \quad . \quad . \qquad (123)$$

for the transforms of the voltage and current at a distance x from the sending end of a transmission line, with sending and receiving end impedances as shown in fig. 11.

The reflexion coefficients are given by (95) and (96). The propagation constant, b, is given by

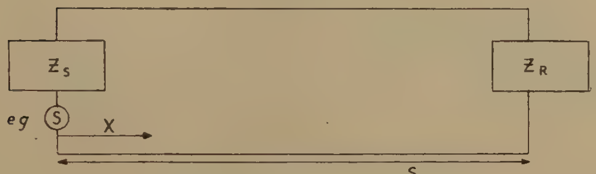
$$b=\sqrt{ZY}. \qquad (124)$$

The Normal Line.

The normal line is characterized by the following parameters :

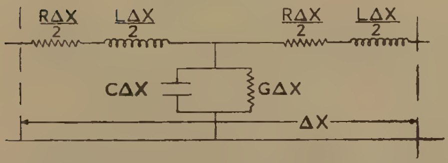
L = the series inductance per unit length.
R = the series resistance per unit length.
G = the shunt conductance per unit length.
C = the shunt capacitance per unit length.

Fig. 11.



Smooth transmission line.

Fig. 12.



$\frac{R\Delta X}{2}$
$\frac{L\Delta X}{2}$
$\frac{R\Delta X}{2}$
$\frac{L\Delta X}{2}$
$C\Delta X$
$G\Delta X$
ΔX

Infinitesimal section of normal line.

Accordingly, an infinitesimal section of the normal line may be represented by fig. 12.

For the normal line, we have

$$Z(p)=Lp+R, \qquad (125)$$

$$Y(p)=Cp+G. \qquad (126)$$

Hence, we have

$$b(p)=\sqrt{(R+Lp)(G+pC)}, \qquad (127)$$

$$Z_0(p)=\sqrt{(R+Lp)/(G+Cp)}, \qquad (128)$$

the operational propagation constant and characteristic impedance of the line respectively.

The Thompson Line.

S. P. Thompson (9), (10) made the proposal that in order to counteract the effect of the line capacity, shunt inductances be introduced between the line conductors. The transient behaviour of the Thompson line has been recently treated by Ekelof (10). An infinitesimal section of the Thompson line is represented by fig. 13.

For the Thompson line, we have

$$Z(p)=Lp+R, \qquad (129)$$

$$Y(p)=Cp+G+1/(R_s+pL_s). \qquad (130)$$

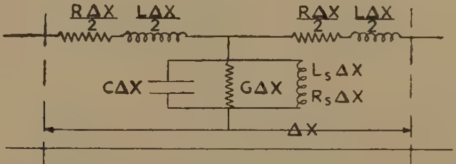
Hence, for the Thompson line, we have

$$b(p)=\sqrt{(R+Lp)(G+pC+1/[R_s+L_s p])}, \qquad (131)$$

$$Z_0(p)=\sqrt{\frac{R+Lp}{G+pC+1/(R_s+L_s p)}}. \qquad (132)$$

The solution for the currents and voltages in the general case is most laborious and follows the same method as that used in determining the response of the cascade connexion of lumped four-terminal networks to an arbitrary impressed electromotive force, as in Section VI.

Fig. 13.



Infinitesimal section of Thompson line.

Expanding (122) in the same manner that (93) was decomposed into forward and backward waves, we obtain

$$E(x,p)=\frac{E_g Z_0}{(Z_s+Z_0)}[e^{-bx}+r_s r_r e^{-b(2s+x)}+(r_s r_r)^2 e^{-b(4s+x)}+]$$
$$+\frac{E_g Z_0 r_r}{(Z_s+Z_0)}[e^{-b(2s-x)}+r_r r_s e^{-b(4s-x)}+\ldots+]. \qquad (133)$$

Similarly, we obtain

$$I(x,p)=\frac{E_g}{(Z_s+Z_0)}[e^{-bx}+r_s r_r e^{-b(2s+x)}+(r_s r_r)^2 e^{-b(4s+x)}+]$$
$$-\frac{E_g r_r}{(Z_s+Z_0)}[e^{-b(2s-x)}+r_r r_s e^{-b(4s-x)}+\ldots]. \qquad (134)$$

In order to obtain the transforms of the expressions (133) and (134) we set up a sequence similar to that of (103), etc., as follows :

$$v_q(t) \doteqdot \frac{E_g Z_0 (r_s r_r)^q}{(Z_s+Z_0)} \quad (q=0,\ 1,\ 2,\ \ldots), \quad . \quad . \quad . \quad . \quad . \quad (135)$$

$$a_x(t) \doteqdot pe^{-bx}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (136)$$

$$w_q(t) \doteqdot \frac{E_g Z_0 (r_s r_r)^q r_r}{(Z_s+Z_0)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (137)$$

Then, by the basic theorem of (106) and (107), we have

$$e(x,\ t) = \sum_{q=0}^{\infty} \int_0^t v_q(t-u) a(u)_{(2qs+x)} \, du$$

$$+ \sum_{q=0}^{\infty} \int_0^t w_q(t-u) a(u)_{[s(2+q)-x]} \, du. \quad . \quad . \quad . \quad (138)$$

Similarly, setting

$$z_q(t) \doteqdot \frac{E_g (r_s r_r)^q}{(Z_s+Z_0)} \quad (q=0,\ 1,\ 2,\ \ldots), \quad . \quad . \quad . \quad . \quad (139)$$

$$y_q(t) \doteqdot \frac{E_g r_r (r_s r_r)}{(Z_s+Z_0)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (140)$$

we obtain

$$i(x,\ t) = \sum_{q=0}^{\infty} \int_0^t z_q(t-u) a(u)_{(2qs+x)} \, du$$

$$- \sum_{q=0}^{\infty} \int_0^t y_q(t-u) a(u)_{[s(2+q)-x]} \, du. \quad . \quad . \quad (141)$$

The solution of the general case of either the normal or the Thompson line presents formidable difficulties. We note that, as in the cascade connexion of four-terminal networks, we have a tremendous simplification if the condition that the receiving end impedance be equal to the characteristic impedance of the line is satisfied. Several investigators (5), (6), (7), (8), (9), (10), (11) have considered many special cases and have evaluated transforms which are most useful in solving many of the special problems of transient behaviour. Since these results are found widely scattered throughout the technical literature, a compilation of these transforms into tabular form was thought to be of value. Many of the transforms given in the following table are found in the extensive table of Fourier Transforms compiled by Campbell and Foster (11), while others are found scattered throughout the literature. By the use of the basic theorems of Table I. and the transforms of Table II. many special problems may be solved in closed form.

Appendix I.

The Laplacian Transformation.

The modern theory of the operational calculus, a mathematical method which has proved to be such a powerful mathematical tool in the study of the transient behaviour of electrical circuits, is based on the following integral, known in the mathematical literature as the Bromwich-Wagner integral :

$$h(t)=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\frac{e^{pt}g(p)\,dp}{p}; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (142)$$

the integration is supposed to be performed in the complex p plane and the integrand is assumed to converge. Following van der Pol (7), (8), we introduce the notation

$$g(p)\doteqdot h(t) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (143)$$

for the relationship between two functions which satisfy the equation (142). Under certain restrictions on the function $g(p)$, the integral (142) possesses the inversion

$$g(p)/p=\int_0^\infty e^{-pt}h(t)\,dt. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (144)$$

This integral is known in the operational literature as Carson's integral. Mathematicians speak of the relation (144) as a Laplacian Transformation of $h(t)$ to the function $g(p)/p$. The relations (142) and (144) taken together are known in the mathematical literature as the Fourier-Mellin integral theorem (13).

The expressions (143) and (144) are valid if either one of the following two groups of conditions, A or B are satisfied :—

A.

1. $g(p)/p$ is an analytic function of the complex variable p, having no points of singularity in any finite region situated to the right of the straight line $Re(p)=a$ parallel to the imaginary axis.

2. There exists a positive number e_0 such that

$$\lim_{\substack{|p|\longrightarrow\infty \\ Re(p)>a}} |p^{\epsilon}g(p)|=0 \quad \text{for} \quad \epsilon<\epsilon_0.$$

B.

1. $g(p)/p$ is an analytic function of the complex variable p, having no points of singularity in any finite region situated to the right of the two straight lines L_1 and L_2 defined by

$$p=a+re^{\pm j\phi} \quad j=\sqrt{-1} \quad \pi>\phi>\pi/2.$$

2. A positive number e_0 exists so that

$$\underset{|p|\longrightarrow\infty}{\text{Limit}}\ |p^e g(p)/p| = 0 \quad \text{for} \quad e < e_0$$

p to the right of L_1 and L_2.

Operators satisfying either the conditions A or B are known as restricted operators. The solution of problems concerning dissipative electrical networks is always expressed by restricted operators, but non-restricted operators sometimes appear as a consequence of evaluating the complete solution by splitting it into parts.

It frequently happens that the inverse of the transform $p^n g(p)$ must be found. That is $h_n(t)$ in the expression

$$h_n(t) \doteqdot p^n g(p), \quad \ldots \ldots \quad (145)$$

where n is a positive integer, $g(p)$ is a restricted operator of class B, but the operator $p^n g(p)$ is a non-restricted operator must be determined. However, it is easy to show that if the integral

$$h(t) = \frac{1}{2\pi j}\int_{L_1 L_2} \frac{e^{pt} g(p)dp}{p}, \quad \ldots \ldots \quad (146)$$

as defined by condition B, converges, then the integral :

$$\frac{d^n h}{dt^n} = \frac{1}{2\pi j}\int_{L_1, L_2} \frac{p^n e^{pt} g(p)dp}{p} \quad \ldots \ldots \quad (147)$$

will converge for $t \geqslant t_0$, where t_0 is positive but arbitrarily small.

Impulse Functions.

Placing $g(p) = 1$ in (146) we obtain

$$h(t) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} \frac{e^{pt}dp}{p}. \quad \ldots \ldots \quad (148)$$

The function $h(t)$, corresponding to the operator $g(p) = 1$, may be shown by the theory of residues to have the following values :

$$h(t) = \begin{Bmatrix} 0 & t<0, \\ 1/2 & t=0, \\ 1 & t>0. \end{Bmatrix}. \quad \ldots \ldots \quad (149)$$

It is thus seen that $h(t)$ is in this case equal to the unit function $1(t)$ of Heaviside, and we may place

$$1(t) \doteqdot 1. \quad \ldots \ldots \quad (150)$$

Now if we place $g(p)=p$ in the integral (146), we obtain the following values:

$$p \doteqdot \begin{Bmatrix} 0 & t<0, \\ \infty & t=0, \\ 0 & t>0. \end{Bmatrix} \quad \text{(151)}$$

Following van der Pol, (7), (8), we can place

$$\delta(t) \doteqdot p, \quad \text{(151 a)}$$

where $\delta(t)$ is the impulse function which has been so extensively used by Dirac in quantum mechanics. The function $\delta(t)$ has the property that

$$\int_{-\infty}^{+\infty} f(t)\delta(t-a)dt = f(a), \quad \text{(152)}$$

where $f(t)$ is a continuous function of t, and

$$\int_{-\infty}^{+\infty} \delta(t)dt = 1. \quad \text{(153)}$$

We may thus treat $\delta(t)$ as an ordinary function having the properties (152) and (153). $\delta(t)$ can be formally differentiated an unlimited number of times, and we write in general

$$p^n \doteqdot \delta^{[n-1]}(t), \quad \text{(154)}$$

where $\delta^{[n-1]}(t)$ is an impulse function of the nth order with the property that

$$\int_{-\infty}^{+\infty} f(t)\delta^{[n]}(t-a)dt = f^{[n]}(a), \quad \text{(155)}$$

and

$$\delta^{[n]}(t) = \begin{Bmatrix} 0 & t \neq 0 \\ \infty & t=0 \end{Bmatrix}. \quad \text{(156)}$$

Table I.

The Basic Theorems of the Operational Calculus.

Starting from the fundamental formulæ (142) and (144), it is possible to deduce some very powerful theorems which are of the utmost importance in the transient analysis of networks. For a proof of these theorems the reader is referred to the work of Carson (5), (6) and van der Pol (7), (8).

If we have two functions satisfying the relations (142) and (144), so that

$$g(p) \doteqdot h(t), \quad \text{(143)}$$

then we have the following theorems :

$$g(p/s) \doteqdot h(st) \quad s>0. \tag{1}$$

$$pg(p)-ph(0) \doteqdot dh/dt. \tag{2}$$

$$p^n g(p)-\sum_{k=0}^{k=n-1} h^{[k]}(0)p^{(n-k)} \doteqdot d^n h(t)/dt^n. \tag{3}$$

$$\int_{-\infty}^{0} h(t)dt+g(p)/p \doteqdot \int_{-\infty}^{t} h(t)dt. \tag{4}$$

$$\frac{p}{p+a}g(p+a) \doteqdot e^{-at}h(t). \tag{5}$$

$$e^{-ap}g(p) \doteqdot \left.\begin{cases} 0, & t<a, \\ h(t-a) & t>a, \end{cases}\right\} a>0. \tag{6}$$

$$e^{ap}g(p) \doteqdot h(t+a) \quad a>0 \quad \text{if} \quad h(t)=0 \quad 0<t<a. \tag{7}$$

$$(-pd/dp)^n g(p) \doteqdot (td/dt)^n h(t) \quad n>0. \tag{8}$$

$$p(-d/dp)^n g(p)/p \doteqdot t^n h(t) \quad n>0. \tag{9}$$

$$\int_p^\infty g(p)dp/p \doteqdot \int_0^t h(u)du/u. \tag{10}$$

$$\int_0^p g(p)dp/p \doteqdot \int_t^\infty h(u)du/u. \tag{11}$$

$$\int_0^\infty g(p)dp/p = \int_0^\infty h(t)dt/t \quad \text{an ordinary equality.} \tag{12}$$

$$\operatorname*{Limit}_{|p|\to\infty} g(p) = \operatorname*{Limit}_{t\to 0} h(t) \quad \text{an ordinary equality.} \tag{13}$$

a. If $g(p)$ has no points of singularity in any finite region situated to the right of the two straight lines L_1, L_2 defined by

$$p=e^{\pm J\phi} \qquad \pi>\phi>\pi/2\,;$$

b. If $g(p)$ tends uniformly towards a finite value $g(0)$ when $|p|\to 0$ on or to the right of the lines L_1 and L_2, then

$$\operatorname*{Limit}_{|p|\to 0} g(p) = \operatorname*{Limit}_{t\to\infty} h(t) \quad \text{an ordinary equality.} \tag{14}$$

$$p\int_p^\infty \int_p^\infty \cdots\cdots \int_p^\infty g(p)(dp)^n/p \doteqdot h(t)/t^n. \tag{15}$$

If $\qquad g_1(p) \doteqdot h_1(t) \quad \text{and} \quad g_2(p) \doteqdot h_2(t)$

then

$$g_1(p)g_2(p)/p \doteqdot \int_0^t h_1(u)h_2(t-u)du$$

$$\doteqdot \int_0^t h_2(u)h_1(t-u)du. \tag{16}$$

If $g(p)=\dfrac{N(p)}{D(p)}$, where $N(p)$ and $D(p)$ are polynomials in p, and the degree of the polynomial $D(p)$ is at least as high as the degree of the polynomial $N(p)$, then, if the polynomial $D(p)$ has the n *distinct* zeros $(p_1, p_2, \ldots p_n)$ and $p=0$ is not a zero of $D(p)$, it follows that

$$\frac{N(p)}{D(p)} \doteq \frac{N(0)}{D(0)} + \sum_{r=1}^{r=n} \frac{N(p_r)e^{p_r t}}{p_r D'(p_r)} \qquad D'(p_r)=dD/dp]_{p=p_r}. \quad (17)$$

If $\qquad g(p)=N(q)/D(q)$

where

a. q is a function of p, $q=q(p)$;

b. $D(q)$ and $N(q)$ are rational polynomials of q;

c. $N(q)$ is of a lower degree than $D(q)$;

d. $D(q)$ has n distinct roots, q_r, and zero is not a root of $D(q)$;

then

$$N(q)/D(q) \doteq N(0)/D(0) + \sum_{r=1}^{r=n} \frac{N(q_r)}{q_r D'(q_r)} b(t, q_r),$$

where

$$\frac{q}{(q-q_r)} \doteq b(t, q_r) \quad \text{and} \quad D'(q_r)=dD/dq]_{q=q_r}. \quad (18)$$

If

$$g(p)=\sum_{r=0}^{\infty} a_r p^r + p^{1/2} \sum_{r=0}^{\infty} b_r p^r,$$

where both series are convergent series in p, then

$$g(p) \doteq a_0 + \frac{1}{t}\left(b_0 - b_1/2t + 1\cdot 3 b_2/(2t)^2 - 1\cdot 3\,.\,5 b_3/(2t)^3 + \ldots\right) \text{ for } t>0. \quad (19)$$

These basic theorems are very powerful when used in connexion with a table of transforms in the solution of transients in networks. Theorem 17 is the conventional Heaviside Expansion formula, while theorem 18 is a generalization of the conventional Heaviside formula. Theorem 19 is known in the operational literature as Heaviside's third rule.

TABLE II.

This table contains a list of transforms of great value for the solution of numerous special cases of the foregoing general theory. These transforms may be obtained by evaluating the basic integrals (142) or (144). The transforms given here are those most frequently occurring in the transient solution of electrical problems, and a collection of these results in one table should be of value.

$$1 \doteq 1(t), \quad \ldots \ldots \ldots \quad (1)$$

where $1(t)$ is the Heaviside "unit function" having the values

$$1(t)=\begin{cases}\infty & t<0,\\ 1/2 & t=0,\\ 1 & t>0.\end{cases}$$

$$p\doteqdot\delta(t), \quad \ldots \ldots \ldots \ldots \quad (2)$$

where $\delta(t)$ is the delta function of Dirac. It satisfies the relations

$$\delta(t)=\begin{cases}\infty & t=0,\\ 0 & t\neq 0.\end{cases}$$

$$\int_{-\infty}\delta(t)dt=1, \qquad \int_{-\infty}^{+\infty}f(t)\delta(t-a)dt=f(a),$$

where $f(t)$ is an integrable continuous function of the real variable t.

$$p^{n}\doteqdot\delta^{[n-1]}(t), \quad \ldots \ldots \ldots \ldots \quad (3)$$

where n is a positive integer; $\delta^{[n-1]}(t)$ is the Dirac impulse function of the nth order and has the properties that

$$\int_{-\infty}^{+\infty}f(t)\delta^{[n]}(t-a)dt=f^{[n]}(a)$$

and

$$\delta^{[n-1]}(t)=\begin{cases}0 & t\neq 0\\ \infty & t=0.\end{cases}$$

$$p^{n}\doteqdot t^{-n}/\Gamma(1-n), \quad \ldots \ldots \ldots \ldots \quad (4)$$

except for n a positive integer, where $\Gamma(n)$ is the Gamma function of the real argument n.

$$p^{1/2}\doteqdot\frac{1}{\sqrt{\pi t}}. \quad \ldots \ldots \ldots \ldots \quad (5)$$

$$p^{3/2}\doteqdot-\sqrt{\frac{2}{\pi}}\frac{1}{(2t)^{3/2}}. \quad \ldots \ldots \ldots \ldots \quad (6)$$

$$p^{5/2}\doteqdot\sqrt{\frac{2}{\pi}}\frac{1.3}{(2t)^{5/2}}. \quad \ldots \ldots \ldots \ldots \quad (7)$$

$$p^{7/2}\doteqdot-\sqrt{\frac{2}{\pi}}\frac{1.3.5}{(2t)^{7/2}}. \quad \ldots \ldots \ldots \ldots \quad (8)$$

$$p^{-1/2}\doteqdot\sqrt{\frac{2}{\pi}}(2t)^{1/2}. \quad \ldots \ldots \ldots \ldots \quad (9)$$

$$p^{-3/2}\doteqdot\sqrt{\frac{2}{\pi}}\frac{(2t)^{3/2}}{1.3}. \quad \ldots \ldots \ldots \ldots \quad (10)$$

$$p^{-5/2} \doteqdot \sqrt{\frac{2}{\pi}}\,\frac{(2t)^{5/2}}{1.3.5}. \tag{11}$$

$$p^{-7/2} \doteqdot \sqrt{\frac{2}{\pi}}\,\frac{(2t)^{7/2}}{1.3.5.7}. \tag{12}$$

$$\frac{p^2}{p+a} \doteqdot \delta(t)-a\epsilon^{-at}. \tag{13}$$

$$\frac{p}{p+a} \doteqdot \epsilon^{-at}. \tag{14}$$

$$\frac{1}{p+a} \doteqdot \frac{1}{a}(1-\epsilon^{-at}). \tag{15}$$

$$\frac{1}{p(p+a)} \doteqdot \frac{t}{a}-\frac{1}{a^2}+\frac{\epsilon^{-at}}{a^2}. \tag{16}$$

$$\frac{p^2}{(p+a)(p+b)} \doteqdot \frac{1}{(a-b)}\{a\epsilon^{-at}-b\epsilon^{-bt}\}. \tag{17}$$

$$\frac{p}{(p+a)(p+b)} \doteqdot \frac{1}{(a-b)}\{\epsilon^{-bt}-\epsilon^{-at}\}. \tag{18}$$

$$\frac{\omega p}{(p^2+\omega^2)} \doteqdot \sin \omega t. \tag{19}$$

$$\frac{p^2}{p^2+\omega^2} \doteqdot \cos \omega t. \tag{20}$$

$$\frac{\omega^2}{p^2+\omega^2} \doteqdot 1-\cos \omega t. \tag{21}$$

$$\frac{\omega p}{p^2-\omega^2} \doteqdot \sinh \omega t. \tag{22}$$

$$\frac{p^2}{p^2-\omega^2} \doteqdot \cosh \omega t. \tag{23}$$

$$\frac{\omega p}{(p+b)^2+\omega^2} \doteqdot \epsilon^{-bt} \sin \omega t. \tag{24}$$

$$\frac{p(p+b)}{(p+b)^2+\omega^2} \doteqdot \epsilon^{-bt} \cos \omega t. \tag{25}$$

$$\frac{\omega p}{(p+b)^2-\omega^2} \doteqdot \epsilon^{-bt} \sinh \omega t. \tag{26}$$

$$\frac{p\omega \cos \phi \pm p^2 \sin \phi}{(p^2+\omega^2)} \doteqdot \sin (\omega t \pm \phi). \tag{27}$$

$$\frac{p^2 \cos \phi \mp \omega p \sin \phi}{(p^2+\omega^2)} \doteqdot \cos (\omega t \pm \phi). \tag{28}$$

$$\frac{\omega p\cos\phi\pm p(p+b)\sin\phi}{(p+b)^2+\omega^2}\doteqdot\epsilon^{-bt}\sin(\omega t\pm\phi). \tag{29}$$

$$\frac{p(p+b)\cos\phi\mp\omega p\sin\phi}{(p+b)^2+\omega^2}\doteqdot\epsilon^{-bt}\cos(\omega t\pm\phi). \tag{30}$$

In the following three transforms, let :

$$\omega^2=\omega_0^{\,2}-a^2,\qquad \tan\phi=\omega/a,$$

$(-m)$ and $(-n)$ be the two roots of

$$p^2+2ap+\omega_0^{\,2}=0. \tag{31}$$

Then

$$\left.\begin{aligned}\frac{p^2}{p^2+2ap+\omega_0^{\,2}}&\doteqdot-\frac{\omega_0}{\omega}\epsilon^{-at}\sin(\omega t-\phi) &&\text{if }\omega_0^{\,2}>a^2,\\ &\doteqdot\frac{1}{n-m}(n\epsilon^{-nt}-m\epsilon^{-mt}) &&\text{if }a^2>\omega_0^{\,2},\\ &\doteqdot\epsilon^{-at}(1-at) &&\text{if }a^2=\omega_0^{\,2}.\end{aligned}\right\} \tag{32}$$

$$\left.\begin{aligned}\frac{p}{p^2+2ap+\omega_0^{\,2}}&\doteqdot\frac{\epsilon^{-at}}{\omega}\sin\omega t &&\text{if }\omega_0^{\,2}>a^2,\\ &\doteqdot\frac{1}{n-m}(\epsilon^{-mt}-\epsilon^{-nt}) &&\text{if }a^2>\omega_0^{\,2},\\ &\doteqdot t\epsilon^{-at} &&\text{if }a^2=\omega_0^{\,2}.\end{aligned}\right. \tag{33}$$

$$\left.\begin{aligned}\frac{1}{p^2+2ap+\omega_0^{\,2}}&\doteqdot\frac{1}{\omega_0^{\,2}}\left\{1-\frac{\omega_0}{\omega}\epsilon^{-at}\sin(\omega t+\phi)\right\} &&\text{if }\omega_0^{\,2}>a^2,\\ &\doteqdot\frac{1}{\omega_0^{\,2}}\left\{1-\frac{\omega_0^{\,2}}{n-m}\left(\frac{\epsilon^{-mt}}{m}-\frac{\epsilon^{-nt}}{n}\right)\right\} &&\text{if }a^2>\omega_0^{\,2},\\ &\doteqdot\frac{1}{\omega_0^{\,2}}\{1-\epsilon^{-at}(1+at)\} &&\text{if }a^2=\omega_0^{\,2}.\end{aligned}\right\}. \tag{34}$$

$$\frac{p^2}{(p+a)^2}\doteqdot\epsilon^{-at}(1-at). \tag{35}$$

$$\frac{p}{(p+a)^2}\doteqdot t\epsilon^{-at}. \tag{36}$$

$$\frac{1}{(p+a)^2}\doteqdot\frac{1}{a^2}\{1-\epsilon^{-at}(1+at)\}. \tag{37}$$

$$\frac{1}{(p+1)^n}\doteqdot\frac{1}{\Gamma(n)}\int_0^t\epsilon^{-u}u^{n-1}\,du, \tag{38}$$

where $n\geqslant 0$.

$$\frac{p}{(p^2+\omega^2)(p+a)}\doteqdot\frac{1}{\omega\sqrt{a^2+\omega^2}}\{\epsilon^{-at}\sin\beta+\sin(\omega t-\beta)\}, \tag{39}$$

where $$\beta = \tan^{-1} \omega/a.$$

$$\frac{p^2+2\omega^2}{p^2+4\omega^2} \doteqdot \cos^2 \omega t. \quad (40)$$

$$\frac{2n\,!}{(p^2+2^2)(p^2+4^2)\ldots\{p^2+(2n)^2\}} \doteqdot \sin^{2n}(t). \quad (41)$$

$$\frac{(2n+1)!\,p}{(p^2+1^2)(p^2+3^2)\ldots\{p^2+(2n+1)^2\}} \doteqdot \sin^{(2n+1)}(t). \quad (42)$$

$$\frac{p}{(p+a)^n} \doteqdot \frac{\epsilon^{-at}t^{n-1}}{(n-1)\,!}. \quad (43)$$

where n is a positive integer.

$$\frac{(p-a)^2}{(p+a)^2} \doteqdot 1-4\,at\,\epsilon^{-at}. \quad (44)$$

$$\frac{p}{\sqrt{p^2+a^2}} \doteqdot \mathrm{J}_0(at), \quad (45)$$

where $\mathrm{J}_0(y)$ is the Bessel function of zeroth order.

$$\frac{p}{\sqrt{p^2-a^2}} \doteqdot \mathrm{J}_0(iat), \quad (46)$$

where $$i=\sqrt{-1}.$$

$$\frac{p}{a^n\sqrt{p^2+a^2}}(\sqrt{p^2+a^2}-p)^n \doteqdot \mathrm{J}_n(at), \quad (47)$$

where $\mathrm{J}_n(y)$ is the Bessel function of order n.

$$\epsilon^{-a/p} \doteqdot \mathrm{J}_0(2\sqrt{at}). \quad (48)$$

$$\left.\begin{aligned} \epsilon^{i/p} &\doteqdot \mathrm{J}_0(2\sqrt{-it}), \\ &\doteqdot \mathrm{ber}\,(2\sqrt{t})+i\,\mathrm{bei}\,(2\sqrt{t}) \end{aligned}\right\}, \quad (49)$$

where $$i=\sqrt{-1}.$$

The ber and bei functions are the "Bessel Real" and "Bessel Imaginary" functions of Lord Kelvin, defined by

$$\mathrm{J}_0(in\sqrt{ix}) = \mathrm{ber}\; nx + i\;\mathrm{bei}\; nx.$$

$$\left.\begin{aligned} \epsilon^{-i/p} &\doteqdot \mathrm{ber}\,(2\sqrt{t})-i\,\mathrm{bei}\,(2\sqrt{t}), \\ &\doteqdot \mathrm{J}_0(2\sqrt{it}) \end{aligned}\right\}. \quad (50)$$

$$\cos(1/p) \doteqdot \mathrm{ber}\,(2\sqrt{t}). \quad (51)$$

$$\sin(1/p) \doteqdot \mathrm{bei}\,(2\sqrt{t}). \quad (52)$$

$$\log_e \frac{1}{\sqrt{1+p^2}} \doteqdot \mathrm{Ci}\,(t), \quad (53)$$

where $$\mathrm{Ci}\,(y)=\int_{-\infty}^{y}\frac{\cos u}{u}\,du,$$

the integral cosine function.

$$\cot^{-1}p\doteqdot\mathrm{Si}\,(t),\qquad\qquad (54)$$

where $$\mathrm{Si}\,(y)=\int_{0}^{y}\frac{\sin u\,du}{u},$$

the integral sine function.

$$\log_e(p-1)\doteqdot-\int_{\infty}^{-t}\frac{\epsilon^{-u}du}{u}.\qquad\qquad (55)$$

$$\sqrt{\frac{p}{p+2a}}\doteqdot\epsilon^{-at}\mathrm{J}_0(iat).\qquad\qquad (56)$$

$$\frac{p}{\sqrt{(p+a)^2-b^2}}\doteqdot\epsilon^{-at}\mathrm{J}_0(ibt).\qquad\qquad (57)$$

$$\epsilon^{-ap}\doteqdot\begin{cases}0 & t<a,\\ \frac{1}{2} & t=a,\quad a>0.\\ 1 & t>a,\end{cases}\qquad\qquad (58)$$

$$\frac{p}{\sqrt{1+p^2}}\epsilon^{-a\sqrt{1+p^2}}\doteqdot\left\{\begin{matrix}0 & t<a,\\ \mathrm{J}_0(\sqrt{t^2-a^2}) & t>a,\end{matrix}\right\}\qquad\qquad (59)$$
$$a>0.$$

$$\frac{p}{\sqrt{p+a}}\doteqdot\frac{\epsilon^{-at}}{\sqrt{\pi t}}.\qquad\qquad (60)$$

$$p\epsilon^{-\sqrt{ap}}\doteqdot\frac{1}{2}\sqrt{\frac{a}{\pi}}\frac{\epsilon^{-a/4t}}{t^{3/2}}.\qquad\qquad (61)$$

$$\epsilon^{-\sqrt{ap}}\doteqdot1-\mathrm{erf}\left(\frac{a}{2\sqrt{t}}\right),\qquad\qquad (62)$$

where erf(y) is the " error function " defined by

$$\mathrm{erf}\,(y)=\frac{2}{\sqrt{\pi}}\int_0^y\epsilon^{-u^2}du.$$

$$\frac{p^{1/2}}{p^{1/2}-a}\doteqdot\epsilon^{a^2t}\{1+\mathrm{erf}\,(at^{1/2})\}.\qquad\qquad (63)$$

$$\frac{p^{1/s}}{p^{1/s}-a}\doteqdot\epsilon^{a^st}\sum_{n=1}^{(s-1)}\{1+\phi_n(t,a)\},\qquad\qquad (64)$$

where $$\phi_n(t,a)=\frac{1}{\Gamma\left(\frac{n}{s}+1\right)}\int_0^{a^nt^{n/s}}\epsilon^{-u^{s/n}}du,\quad s>0.$$

$$\frac{p^3}{p^3-a}\doteqdot\left(\frac{1}{3}\right)\epsilon^{a^{1/3}t}+\left(\frac{2}{3}\right)\epsilon^{-a^{1/3}t/2}\cos\left(a^{1/3}t\frac{\sqrt{3}}{2}\right).\qquad\qquad (65)$$

$$\frac{p^4}{p^4-a} \doteqdot \frac{1}{2}\cosh(a^{1/4}t)+\frac{1}{2}\cos(a^{1/4}t). \qquad (66)$$

$$\frac{q}{q-a} \doteqdot \frac{b}{b-a^2}-\frac{a^2}{b-a^2}\epsilon^{(a^2-b)t}$$

$$+\frac{a\sqrt{b}}{b-a^2}\operatorname{erf}(\sqrt{bt})-\frac{a^2}{b-a^2}\epsilon^{(a^2-b)t}\operatorname{erf}(a\sqrt{t}), \qquad (67)$$

where $q=(p+b)^{1/2}$ and $b\neq a^2$.

$$\frac{q}{q-a} \doteqdot \frac{1}{(1-a^2)}\left\{\epsilon^{a^2bt/(1-a^2)}+a\epsilon^{-(bt/2)}\mathrm{I}_0\left(\frac{bt}{2}\right)\right.$$

$$\left.+\frac{ab}{(1-a^2)^2}\epsilon^{a^2bt/(1-a^2)}\int_0^t \epsilon^{(a^2+1)bt/2(a^2-1)}\mathrm{I}_0\left(\frac{bt}{2}\right)dt\right\}, \qquad (68)$$

where $q=\left(\frac{p}{p+b}\right)^{1/2}$ $a^2\neq 1$.

$\mathrm{I}_0(y)=\mathrm{J}_0(iy)=$the Bessel function of the first kind.

$$\frac{p^{3/2}}{(p+a)} \doteqdot \frac{1}{\sqrt{\pi t}}+ia^{1/2}\epsilon^{-at}\operatorname{erf}(ia^{1/2}t^{1/2}), \qquad (69)$$

where $i=\sqrt{-1}$.

$$\frac{1}{p^4-3p^2+2} \doteqdot \frac{1}{2}+\frac{1}{2}\cosh t\sqrt{2}-\cosh t. \qquad (70)$$

$$\frac{\omega p^{3/2}}{(p^2+\omega^2)} \doteqdot \sqrt{2\omega}\{\sin(\omega t)\,\mathrm{S}(\omega^{1/2}t^{1/2}\sqrt{2/\pi})$$

$$+\cos(\omega t)\mathrm{C}(\omega^{1/2}t^{1/2}\sqrt{2/\pi}), \qquad (71)$$

where $\mathrm{C}(y)$ and $\mathrm{S}(y)$ are the Fresnel integrals defined by

$$\mathrm{C}(y)=\int_0^y \cos\left(\frac{\pi u^2}{2}\right)du,$$

$$\mathrm{S}(y)=\int_0^y \sin\left(\frac{\pi u^2}{2}\right)du.$$

$$\frac{\sinh(bp^{1(2)})}{\sinh(ap^{1/2})} \doteqdot \frac{b}{a}+\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\sin\left(\frac{n\pi b}{a}\right)\epsilon^{-(n^2\pi^2/a^2)t}. \qquad (72)$$

$$\frac{\sinh b(p+c)^{1/2}}{\sinh a(p+c)^{1/2}} \doteqdot \frac{b}{a}$$

$$+\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\sin\left(\frac{n\pi b}{a}\right)\left\{\frac{c}{c+\frac{n^2\pi}{a^2}}+\frac{n^2\pi^2}{a^2c+n^2\pi^2}\epsilon^{-(n^2\pi^2/a^2+c)t}\right\}. \qquad (73)$$

$$\frac{\cosh(x\sqrt{\mathrm{RC}p})}{\cosh(l\sqrt{\mathrm{RC}p})} \doteqdot 1+\frac{4}{\pi}\sum_{s=1}^{\infty}\frac{(-1)^s}{(2s-1)}\cos\left\{\left(\frac{2s-1}{2}\right)\frac{\pi x}{l}\right\}\epsilon^{-\frac{(2s-1)^2\pi^2t}{4c\mathrm{R}l^2}}. \qquad (74)$$

$$\frac{(p\mathrm{C}^{1/2}\sinh(x\sqrt{\mathrm{RC}p})}{\sqrt{\mathrm{R}}\cosh(l\sqrt{\mathrm{RC}p})} \doteqdot -\frac{2}{l\mathrm{R}}\sum_{s=1}^{\infty}(-1)^s\sin(m_s x)\epsilon^{-\frac{m_s^2 t}{\mathrm{CR}}} \quad . \quad . \quad . \quad . \quad . \quad . \qquad (75)$$

$$m_s=\left(\frac{2s-1}{2}\right)\frac{\pi}{l}.$$

$$\frac{(p\mathrm{C})^{1/2}}{\sqrt{\mathrm{R}}}\,\frac{\cosh(x\sqrt{\mathrm{RC}p})}{\sinh(l\sqrt{\mathrm{RC}p})} \doteqdot \frac{1}{\mathrm{R}l}+\frac{2}{\mathrm{R}l}\sum_{s=1}^{\infty}(-1)^s\epsilon^{-\frac{s^2\pi^2 t}{\mathrm{CR}l^2}}\cos\left(\frac{s\pi x}{l}\right). \qquad (76)$$

If $\alpha=\sqrt{(\mathrm{R}+\mathrm{L}p)(\mathrm{G}+p\mathrm{C})}$

$\mathrm{Z}=\mathrm{R}+\mathrm{L}p$

$$\frac{\alpha\cosh\alpha x}{\mathrm{Z}\sinh\alpha l} \doteqdot \sqrt{\frac{\mathrm{G}}{\mathrm{R}}}\,\frac{\cosh x\sqrt{\mathrm{RG}}}{\sinh l\sqrt{\mathrm{RG}}}-\frac{\mathrm{E}}{\mathrm{R}l}\epsilon^{-\frac{\mathrm{R}}{\mathrm{L}}t}$$

$$+2\sum_{s=1}^{\infty}\frac{(-1)^s l\cos\left(\frac{s\pi x}{l}\right)}{s^2\pi^2+\mathrm{RG}l^2}\epsilon^{-\rho t}\left[\frac{1}{\beta}\left\{\frac{s^2\pi^2}{\mathrm{L}l^2}+\frac{1}{2}\mathrm{G}\left(\frac{\mathrm{R}}{\mathbf{L}}-\frac{\mathrm{G}}{l}\right)\right\}\right.$$

$$\left.\times\sin\beta_s t-\mathrm{G}\cos\beta_s t]\right], \qquad (77)$$

where
$$\rho=\frac{1}{2}\left(\frac{\mathrm{R}}{\mathrm{L}}+\frac{\mathrm{G}}{\mathrm{C}}\right) \quad m_s=\frac{s\pi}{l};$$

$$\beta_s=\sqrt{\frac{m_s^2}{\mathrm{LC}}-\frac{1}{4}\left(\frac{\mathrm{R}}{\mathrm{L}}-\frac{\mathrm{G}}{\mathrm{C}}\right)^2}.$$

$$\frac{\cosh\alpha(l-x)}{\cosh\alpha l} \doteqdot \frac{\cosh s(l-x)}{\cosh sl}$$

$$-\frac{\pi v^2}{l^2}\epsilon^{\rho t}\sum_{n=1,3,5}^{\infty} n\sin\frac{n\pi x}{2l}\,\frac{(\rho\sin\beta_n t+\beta_n\cos\beta_n t)}{\beta_n(\rho^2+\beta_n)}, \qquad (78)$$

where
$$\beta_n=\sqrt{\frac{n^2\pi^2v^2}{4l^2}-\sigma^2}, \qquad \rho=\frac{1}{2}\left(\frac{\mathrm{R}}{\mathrm{L}}+\frac{\mathrm{G}}{\mathrm{C}}\right),$$

$$\alpha=\sqrt{(\mathrm{R}+\mathrm{L}p)(\mathrm{G}+p\mathrm{C})}, \qquad \sigma=\frac{1}{2}\left(\frac{\mathrm{R}}{\mathbf{L}}-\frac{\mathrm{G}}{\mathrm{C}}\right),$$

$$s=\sqrt{\mathrm{RG}}, \qquad v=\frac{1}{\sqrt{\mathrm{LC}}}.$$

$$\frac{\sinh\alpha(l-x)}{\mathrm{Z}_0\cosh\alpha l} \doteqdot \frac{\mathrm{G}}{s}\,\frac{\sinh s(l-x)}{\cosh sl}$$

$$+\frac{2v^2}{l}\epsilon^{-\rho t}\sum_{n=1,3,5}^{\infty}\cos\frac{n\pi x}{2l}\left\{\frac{\mathrm{G}\rho\sin\beta_n t+\mathrm{G}\beta_n\cos\beta_n t}{\beta_n(\rho^2+\beta_n^2)}-\frac{\mathrm{C}\sin\beta_n t}{\beta_n}\right\}, \qquad (79)$$

where the constants are those defined for (78), and

$$\mathrm{Z}_0=\sqrt{\frac{\mathrm{R}+\mathrm{L}p}{\mathrm{G}+\mathrm{C}p}}.$$

$$\frac{\sinh \alpha(l-x)}{\sinh \alpha l} \doteqdot \frac{\sinh s(l-x)}{\sinh sl}$$

$$-\frac{2v^2\pi}{l^2}\epsilon^{-\rho t} \sum_{n=1,2,3}^{\infty} \frac{n \sin \frac{n\pi x}{l}\{\rho \sin \beta_n t+\beta_n \cos \beta_n t\}}{\beta_n(\rho^2+\beta_n{}^2)}, \qquad (80)$$

where the constants are those defined in (78).

$$\epsilon^{-\alpha x} \doteqdot \left\{ \begin{matrix} 0 & t<x/v \\ \epsilon^{-\frac{\rho x}{v}}+\frac{\sigma x}{v}\int_{x/v}^{t} \epsilon^{-\rho u} \frac{\mathrm{I}_1(\sigma\sqrt{u^2-x^2/r^2})}{\sqrt{u^{\rho}-x^2/r^2}}\,du & t>x/v \end{matrix} \right\}, \qquad (81)$$

where $\alpha=\sqrt{(\mathrm{L}p+\mathrm{R})(\mathrm{C}p+\mathrm{G})}$, $\rho=\frac{1}{2}\left(\frac{\mathrm{R}}{\mathrm{L}}+\frac{\mathrm{G}}{\mathrm{C}}\right)$,

$\sigma=\frac{1}{2}\left(\frac{\mathrm{R}}{\mathrm{L}}-\frac{\mathrm{G}}{\mathrm{C}}\right)$, $v=\frac{1}{\sqrt{\mathrm{LC}}}$.

$\mathrm{I}_1(y)$ is the Bessel function of the first kind.

$$\frac{p \exp\left\{-\frac{x}{v}\sqrt{(p+\rho)^2-\sigma^2}\right\}}{\sqrt{(p+\rho)^2-\sigma^2}} \doteqdot \left\{ \begin{matrix} 0 & t<x/v \\ \epsilon^{-\rho t}\mathrm{I}_0(\sigma\sqrt{t^2-x^2/v^2}) & t>x/v. \end{matrix} \right\}. \qquad (82)$$

$$p\epsilon^{-\alpha x} \doteqdot \delta\left(t-\frac{x}{v}\right)\epsilon^{-\rho x/v}+\frac{\sigma x}{vz}\epsilon^{-\rho t}\mathrm{I}_1(\sigma z) \quad t>x/v, \quad . \quad . \quad . \qquad (83)$$

where $\alpha=\sqrt{(\mathrm{L}p+\mathrm{R})(\mathrm{C}p+\mathrm{G})}$, $\mathrm{Z}=\sqrt{t^2-(x/v)^2}$,

$v=\frac{1}{\sqrt{\mathrm{LC}}}$, $\rho=\frac{1}{2}\left(\frac{\mathrm{R}}{\mathrm{L}}+\frac{\mathrm{G}}{\mathrm{C}}\right)$,

$\sigma=\frac{1}{2}\left(\frac{\mathrm{R}}{\mathrm{L}}-\frac{\mathrm{G}}{\mathrm{C}}\right)$.

$$\frac{p\epsilon^{-\alpha x}}{z_0} \doteqdot \left\{ \begin{matrix} 0 & t<x/v \\ \frac{1}{k}\epsilon^{-\frac{\rho x}{v}}\delta\left(t-\frac{x}{v}\right)+\frac{1}{k}\epsilon^{-\rho t}\left[\frac{\sigma t}{z}\mathrm{I}_1(\sigma z)-\sigma\mathrm{I}_0(\sigma z)\right] & t>x/v \end{matrix} \right\}, \qquad (84)$$

where $\alpha=\sqrt{(\mathrm{L}p+\mathrm{R})(\mathrm{C}p+\mathrm{G})}$, $\mathrm{Z}=\sqrt{t^2-\left(\frac{x}{v}\right)^2}$,

$k=\sqrt{\frac{\mathrm{L}}{\mathrm{C}}}$, $\sigma=\frac{1}{2}\left(\frac{\mathrm{R}}{\mathrm{L}}-\frac{\mathrm{G}}{\mathrm{C}}\right)$,

$v=\frac{1}{\sqrt{\mathrm{LC}}}$, $\rho=\frac{1}{2}\left(\frac{\mathrm{R}}{\mathrm{L}}-\frac{\mathrm{G}}{\mathrm{C}}\right)$,

$\mathrm{Z}_0=\sqrt{\frac{\mathrm{L}p+\mathrm{R}}{\mathrm{C}p+\mathrm{G}}}$, $\mathrm{Z}=\sqrt{t^2-\left(\frac{x}{v}\right)^2}$.

$$p\epsilon^{-\alpha x} \doteqdot \frac{y}{2t\sqrt{\pi t}}\exp\left(-\frac{y^2}{4t}-2\beta t\right) \quad t>0. \quad . \quad . \quad . \quad . \qquad (85)$$

$$\alpha=\sqrt{\mathrm{R}(\mathrm{C}p+\mathrm{G}}, \qquad \beta=\frac{\mathrm{G}}{2\mathrm{C}}, \qquad y=x\sqrt{\mathrm{RC}}.$$

$$\frac{p\epsilon^{-\alpha x}}{\mathrm{Z}_0}\doteqdot\frac{u(y^2-2t}{4t^2\sqrt{\pi t}}\exp\left(-\frac{y^2}{4t}-2\beta t\right) \quad t>0, \quad . \quad . \quad . \quad (86)$$

where $\alpha=\sqrt{\mathrm{R}(\mathrm{C}p+\mathrm{G})}, \qquad u=\sqrt{\mathrm{C}/\mathrm{R}}, \qquad y=x\sqrt{\mathrm{RC}},$

$$\mathrm{Z}_0=\sqrt{\frac{\mathrm{R}}{\mathrm{G}+\mathrm{C}p}}.$$

$$\epsilon^{-\alpha x}\doteqdot\frac{1}{2}\left[\epsilon^{-y\sqrt{2\beta}}\operatorname{erf}\left(\frac{y}{2\sqrt{t}}-\sqrt{2\beta t}\right)+\epsilon^{y\sqrt{2\beta}}\operatorname{erf}\left(\frac{y}{2\sqrt{t}}+\sqrt{2\beta t}\right)\right] \quad t>0, \quad . \quad . \quad . \quad (87)$$

where $\alpha=\sqrt{\mathrm{R}(\mathrm{C}p+\mathrm{G})}, \qquad y=x\sqrt{\mathrm{RC}}, \qquad \beta=\frac{\mathrm{G}}{2\mathrm{C}},$

$$y=x\sqrt{\mathrm{RC}}, \qquad \operatorname{erf}(y)=\frac{2}{\sqrt{\pi}}\int_0^y\epsilon^{-u^2}du.$$

$$\frac{\epsilon^{-\alpha x}}{\mathrm{Z}_0}\doteqdot\frac{u}{\sqrt{\pi t}}\exp\left(-\frac{y^2}{4t}-2\beta t\right)$$
$$+\frac{u\sqrt{2\beta}}{2}\left[\epsilon^{-y\sqrt{2\beta}}\operatorname{erf}\left(\frac{y}{2\sqrt{t}}-\sqrt{2\beta t}\right)-\epsilon^{y\sqrt{2\beta}}\operatorname{erf}\left(\frac{y}{2\sqrt{t}}+\sqrt{2\beta t}\right)\right]$$
$$t>0, \qquad (88)$$

where $\alpha=\sqrt{\mathrm{R}(\mathrm{C}p+\mathrm{G})}, \qquad \beta=\frac{\mathrm{G}}{2\mathrm{C}}, \qquad u=\sqrt{\mathrm{C}/\mathrm{R}},$

$$\mathrm{Z}_0=\sqrt{\frac{\mathrm{R}}{\mathrm{C}p+\mathrm{G}}}, \qquad y=x\sqrt{\mathrm{RC}}.$$

$$\left.\begin{aligned}\frac{p\epsilon^{-\alpha x}}{\mathrm{Z}_0}&\doteqdot 0 \quad t<x/v,\\ &\doteqdot\frac{\delta}{k}(t-x/v)\epsilon^{\frac{\alpha x}{v}}+\frac{1}{k}\epsilon^{-at}\left[\frac{at}{z}\mathrm{I}_1(az)-a\mathrm{I}_0(az)\right] \quad t>x/v\end{aligned}\right\}, \quad . \quad (89)$$

where $\alpha=\sqrt{(\mathrm{L}p+\mathrm{R})\mathrm{C}p}, \qquad k=(\mathrm{L}/\mathrm{C})^{1/2}, \qquad \mathrm{Z}=\sqrt{t^2-(x/v)^2},$

$$\mathrm{Z}_0=\sqrt{\frac{\mathrm{L}p+\mathrm{R}}{\mathrm{C}p}}, \qquad v=(\mathrm{L}/\mathrm{C})^{-1/2}, \qquad a=\frac{\mathrm{R}}{2\mathrm{L}}.$$

$$\left.\begin{aligned}\frac{\epsilon^{\alpha x}}{\mathrm{Z}_0}&\doteqdot 0 \quad t<x/v,\\ &\doteqdot\frac{1}{k}\epsilon^{-at}\mathrm{I}_0(a\mathrm{Z}) \quad t>x/v\end{aligned}\right\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (90)$$

where the constants have the same value as in (89).

$$p\epsilon^{-\alpha x}\doteqdot\delta(t-x/v)\epsilon^{-\frac{\rho x}{v}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (91)$$

where $$\alpha=\sqrt{(\mathrm{L}p+\mathrm{R})(\mathrm{G}+\mathrm{C}p)} \quad \text{and} \quad \frac{\mathrm{R}}{\mathrm{L}}=\frac{\mathrm{G}}{\mathrm{C}},$$

$$\rho=\frac{\mathrm{R}}{2\mathrm{L}}+\frac{\mathrm{G}}{2\mathrm{C}}.$$

$$\left.\begin{aligned}\epsilon^{-\alpha x}&\doteqdot 0 && t<x/v,\\ &\doteqdot \epsilon^{-\frac{\rho x}{v}} && t>x/v.\end{aligned}\right\}, \quad \ldots \ldots \quad (92)$$

where the constants have the values given in (91).

$$\frac{p\sqrt{p+2a}}{\sqrt{p}+\sqrt{p+2a}}\doteqdot\frac{\delta(t)}{2}+\frac{1}{2t}\epsilon^{-at}\mathrm{I}_1(at) \quad t>0, \quad \ldots \quad (93)$$

where $$a=\frac{\mathrm{R}}{2\mathrm{L}}.$$

$$\frac{\sqrt{p+2a}}{\sqrt{p}+\sqrt{p+2a}}\doteqdot 1-\frac{1}{2}\epsilon^{-at}\{\mathrm{I}_0(at)+\mathrm{I}_1(at)\} \quad t>0. \quad \ldots \quad (94)$$

$$\frac{p\exp(-y\sqrt{p})}{1+\sqrt{p/a}}\doteqdot\sqrt{\frac{a}{\pi t}}\exp\left(\frac{-y^2}{4t}\right)$$
$$-a\exp(y\sqrt{a}+at)\operatorname{erf}\left(\frac{y}{2\sqrt{t}}+\sqrt{at}\right) \quad t>0. \quad (95)$$

$$\frac{\exp(-y\sqrt{p})}{1+\sqrt{p/a}}\doteqdot\operatorname{erf}\frac{y}{2\sqrt{t}}-\exp(y\sqrt{a}+at)\operatorname{erf}\left(\frac{y}{2\sqrt{t}}+\sqrt{at}\right)$$
$$t>0. \quad \ldots \ldots \quad (96)$$

$$\frac{pu\sqrt{p}\exp(-y\sqrt{p})}{1+\sqrt{p/a}}\doteqdot\frac{u(y-2t\sqrt{a})}{2t}\sqrt{\frac{a}{\pi t}}\exp\left(\frac{-y^2}{4t}\right)$$
$$+ua\sqrt{a}\exp(y\sqrt{a}+at)\operatorname{erf}\left(\frac{y}{2\sqrt{t}}+\sqrt{at}\right)$$
$$t>0. \quad \ldots \ldots \quad (97)$$

$$\frac{u\sqrt{p}\exp(-y\sqrt{p})}{1+\sqrt{p/a}}\doteqdot u\exp(y\sqrt{a}+at)\operatorname{erf}\left(\frac{y}{2\sqrt{t}}+\sqrt{at}\right),$$
$$t>0. \quad \ldots \ldots \quad (98)$$

$$\frac{pu\sqrt{p}}{1+\sqrt{p/a}}\doteqdot u\sqrt{a}\left\{\delta(t)-\sqrt{\frac{a}{\pi t}}+a\epsilon^{at}\operatorname{erf}\sqrt{at}\right. \quad t>0. \quad \ldots \quad (99)$$

$$\frac{u\sqrt{p}}{1+\sqrt{p/a}}\doteqdot u\sqrt{a}\epsilon^{at}\operatorname{erf}\sqrt{at}. \quad \ldots \ldots \quad (100)$$

$$\frac{p^2\,\mathrm{cxp}\,(-y\sqrt{p})}{1+\sqrt{p/a}} \doteqdot \frac{(y^2-2yt\sqrt{a}-2t+4at^2}{4t^2}\sqrt{\frac{a}{\pi t}}\exp\left(\frac{-y^2}{4t}\right)$$

$$-a^2\exp(y\sqrt{a}+at)\,\mathrm{erf}\left(\frac{y}{2\sqrt{t}}+\sqrt{at}\right) \quad t>0. \quad . \quad (101)$$

$$\frac{p\exp\,(-y\sqrt{p})}{1+\sqrt{p/a}} \doteqdot \sqrt{\frac{a}{\pi t}}\exp\left(\frac{-y^2}{4t}\right)$$

$$-a\exp\,(y\sqrt{a}+at)\,\mathrm{erf}\left(\frac{y}{2\sqrt{t}}-\sqrt{at}\right) \quad t>0. \quad (102)$$

$$\frac{p^2}{1+\sqrt{p/a}} \doteqdot -a\delta(t)\frac{(2at-1)}{2t}\sqrt{\frac{a}{\pi t}}-a^2\epsilon^{at}\,\mathrm{erf}\,\sqrt{at} \quad t>0. \quad (103)$$

$$\frac{p}{1+\sqrt{p/a}} \doteqdot \sqrt{\frac{a}{\pi t}}-a\epsilon^{at}\,\mathrm{erf}\,(\sqrt{at}) \quad t>0 \quad . \quad . \quad . \quad . \quad . \quad (104)$$

$$\frac{pw^{(2n+1)}\{\sqrt{(p+a)^2+w^2}+(p+a)\}^{-2n}}{k\sqrt{(p+a)^2+w^2}} \doteqdot \frac{2}{\mathrm{L}}\epsilon^{-at}\mathrm{J}_{2n}(wt), \quad . \quad (105)$$

where $k=(\mathrm{L}/\mathrm{C})^{1/2}, \qquad a=\frac{\mathrm{R}}{\mathrm{L}}=\frac{\mathrm{G}}{\mathrm{C}},$

$w=2(\mathrm{LC})^{-1/2}$, n is a positive integer.

$$\frac{p2(2a)^n}{\mathrm{R}}\sqrt{\frac{p}{p+2a}}(\sqrt{p+2a}+\sqrt{p})^{-2n}$$

$$\doteqdot \frac{a}{\mathrm{R}}\epsilon^{-at}\{\mathrm{I}_{n-1}(at)-2\mathrm{I}_n(at)+\mathrm{I}_{n+1}(at)\} \quad t>0, \quad (106)$$

where $a=\frac{2}{\mathrm{RC}}$, n is a positive integer.

$\mathrm{I}_n(y)$ is the Bessel function of the first kind and the nth order.

$$\frac{2(2a)^n}{\mathrm{R}}\sqrt{\frac{p}{p+2a}}(\sqrt{p+2a}+\sqrt{p})^{-2n} \doteqdot \frac{2}{\mathrm{R}}\epsilon^{-at}\mathrm{I}_n(at) \quad t>0^{-2n}, \quad . \quad (107)$$

where the constants are the same as in 106

$$\frac{\{\sqrt{1+a}+\sqrt{a}\}^{-2n}}{\sqrt{(1+a)\mathrm{Z}_1\mathrm{Z}_2}} \doteqdot \frac{1}{k}\int_0^{\omega_c t}\epsilon^{-\frac{b}{\omega c}u}\,\mathrm{J}_{2n}(u)du, \quad . \quad . \quad . \quad . \quad (108)$$

where n is a positive integer.

$$a=\frac{\mathrm{LC}}{4}(p+b)^2, \qquad \frac{1}{\mathrm{Z}_2}=\mathrm{C}(p+b), \qquad k=(\mathrm{L}/\mathrm{C})^{1/2}.$$

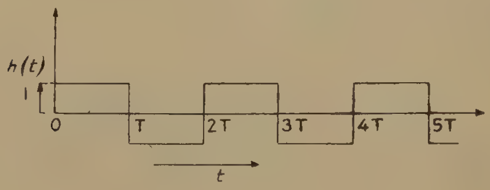
$$\mathrm{Z}_1=\mathrm{L}(p+b), \qquad \omega_c=\frac{2}{\sqrt{\mathrm{LC}}}$$

$$\frac{\omega_c}{k}\,\frac{1}{\sqrt{p^2+\omega_c^2}}\left(\frac{\sqrt{p^2+\omega_c^2}-p}{\omega_c}\right)^{2n} \doteqdot \frac{1}{k}\int_0^{\omega_c t}\mathrm{J}_{2n}(u)du, \quad . \quad . \quad (109)$$

where n is a positive integer.

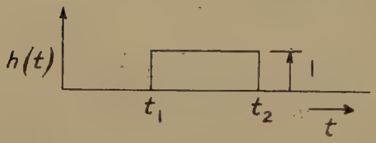
$$\tfrac{1}{2}\tanh\frac{\mathrm{TP}}{2} \doteqdot h(t), \qquad (110)$$

where


h(t)
1
0
T
2T
3T
4T
5T
t

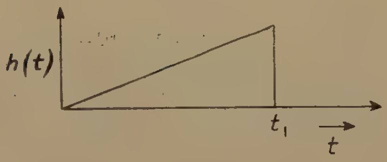
$$\epsilon^{-t_1p}-\epsilon^{-t_2p} \doteqdot h(t) \quad \begin{matrix} t_1>0 \\ t_2>t_1 \end{matrix}, \qquad (111)$$

where



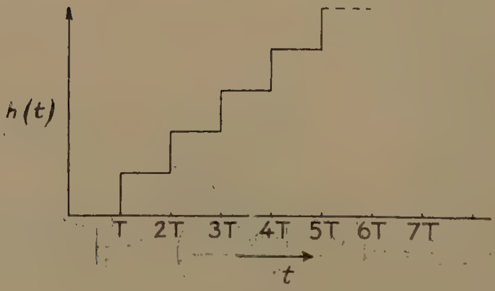
$$\frac{m}{p}(1-\epsilon^{-pt_1})-mt_1\epsilon^{-pt_1} \doteqdot h(t), \qquad (112)$$

where slope $=m$.



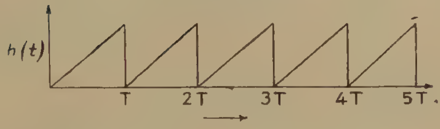
$$\frac{1}{(\epsilon^{\mathrm{T}p}-1)} \doteqdot h(t), \qquad (113)$$

where


h(t)
T
2T
3T
4T
5T
6T
7T
t

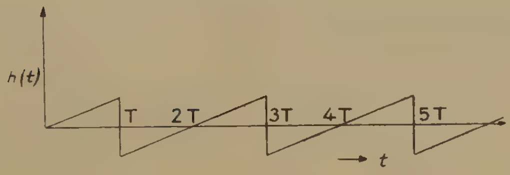
$$\frac{m}{p}-\frac{mT}{\epsilon^{pT}-1}\doteqdot h(t), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (114)$$

where slope $=m$.



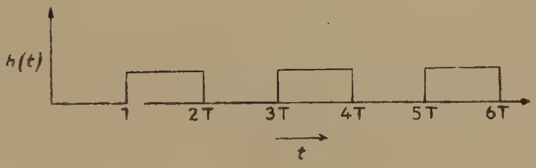
$$\frac{m}{p}-2mT\left\{\frac{1}{\epsilon^{Tp}-1}-\frac{1}{\epsilon^{2Tp}-1}\right\}\doteqdot h(t), \quad . \quad . \quad . \quad . \quad . \qquad (115)$$

where slope $=m$.



$$\frac{1}{\epsilon^{Tp}1+1}\doteqdot h(t), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (116)$$

where :



APPENDIX II.

Illustrative Example.

The use of the transforms of Table II. and the basic theorems of Table I. makes it possible to solve a large class of problems of transients in four terminal networks in closed form. As an illustration, let us consider the problem of determining the current flowing in the nth section of a semi-infinite artificial line as shown in fig. 14, when a steady electromotive force of magnitude E is impressed across the sending end terminal at $t=0$. The line is one with mid-series termination.

This line may be regarded as a cascade connexion of identical four-terminal networks of the T-type, as shown in fig. 15.

By the theory of part V, C, it is easy to show that the associated matrix of this structure is given by

$$[u]=\begin{bmatrix}(1+Z_1/2Z_2), & (Z_1/2+Z_1^2/4Z_2+Z_1/2), \\ 1/Z_2, & (1+Z_1/2Z_2)\end{bmatrix} \quad . \; . \; . \; . \; . \; . \quad (157)$$

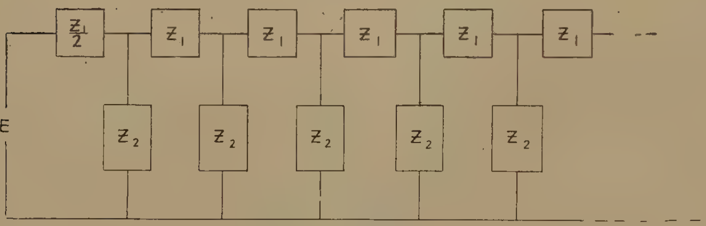
If we let

$$b=Z_1/4Z_2, \quad . \; . \; . \; . \; . \; . \; . \; . \; . \quad (158)$$

then, by (61), we obtain

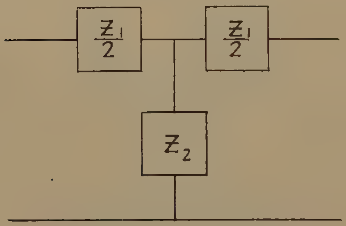
$$m_{1,\,2}=(1+2b)\pm\sqrt{(1+2b)^2-1}, \quad . \; . \; . \; . \; . \; . \quad (159)$$

Fig. 14.



Artificial line.

Fig. 15.



and hence :

$$\begin{aligned} e^a &=(1+2b)+\sqrt{(1+2b)^2-1} \\ &=(\sqrt{1+b}+\sqrt{b})^2 \quad . \; . \; . \; . \; . \; . \; . \; . \; . \quad (160)\end{aligned}$$

Now equation (94) for the transform of the current in the *k*th section of the network reduces to

$$I_k(p)=E_g e^{-ka}/Z_{01} \quad . \; . \; . \; . \; . \; . \; . \; . \quad (161)$$

for the case of a semi-infinite artificial line having a voltage whose transform is $E_g(p)$ directly impressed on it.

By (78) we have

$$Z_{01}=\frac{1}{\sqrt{Z_1Z_2}}\frac{1}{\sqrt{1+b}}. \qquad (162)$$

If the voltage impressed on the line at $t=0$ is a constant voltage of magnitude E, we have

$$E_g(p)=E. \qquad (163)$$

Hence, substituting (160), (162), and (163) into (161), we obtain

$$I_k(p)=\frac{E(1+b+b)^{-2}}{\sqrt{Z_1Z_2}\sqrt{1+b}}. \qquad (164)$$

The nature of the inverse transform of $I_k(p)$ depends upon the character of the impedances Z_1 and Z_2. Let us, for simplicity, discuss the quasi-distortionless case of the low pass filter. This case is characterized by having a series inductance L and resistance R, and a shunt leakage G and capacitance C, so that

$$Z_1=Lp+R, \qquad (165)$$

$$Z_2=Cp+G, \qquad (166)$$

where

$$b=R/L=G/C. \qquad (167)$$

For this case, we have

$$b=LC(p+b)^2/4. \qquad (168)$$

Substituting into (164) we see that we may make use of transform ♯108 of Table II. if we place $a=b$. We thus obtain

$$i_k(t)=\frac{E}{s}\int_0^{w_ct}e^{-bu/w_c}J_{2k}(u)du, \qquad (169)$$

where

$$w_c=2/\sqrt{LC}, \qquad (170)$$

and

$$s=\sqrt{L/C} \qquad (171)$$

for the current of the *ktl* mesh of the structure of fig. 14, subject to the restrictions of (165), (166), and (167), when a steady electromotive force of magnitude E is impressed upon the system.

This example illustrates the power and simplicity of the solution of transients in four-terminal networks, using the powerful tools of the Laplacian transformation and matrix algebra. If the particular case under consideration leads to a transform not found in Table II., then by the application of the basic theorems of Table I. it is sometimes possible to obtain the inverse transform by obtaining forms that are found in Table II. For more complicated transforms the larger table of Fourier transforms, compiled by Campbell and Foster $^{(11)}$, may be consulted.

Bibliography.

(1) Streker and Feldtkeller, " Grundlagen der Theory des Allgemeinen Vierpols," *Elektrische Nachrichten Technik*, p. 93 (1929).
(2) Howitt, "Group Theory and the Electric Circuit," Physical Review, p. 1583 (1931).
(3) Brillouin, " Les Filtres Electriques et la Theorie des Matrices," ' Revue Generale de l'Electricite,' p. 3 (1936).
(4) Pipes, " Matrices in Engineering," Electrical Engineering, p. 1177 (1937).
(5) J. R. Carson, " Theory of the Transient Oscillations of Electrical Networks and Transmission Systems," J. Am. Inst. Elec. Eng. xxxviii. pp. 345–427 (1919).
(6) J. R. Carson and O. J. Zobel, " Transient Oscillations in Electric Wave Filters," B. S. T. J. (June 1923).
(7) van der Pol, " On the Operational Solution of Linear Differential Equations and an Investigation of the Properties of these Solutions," Phil. Mag. viii. p. 861 (1929).
(8) van der Pol and Niessen, "Symbolic Calculus," Phil. Mag. xiii. p. 537 (1932).
(9) Stig Ekelof, " The Transients of an Inductively Shunted Electric Transmission Line.—Part I." Ericsson Technics, no. 5 (1937).
(10) Stig Ekelof, " The Transients of an Inductively Shunted Electric Transmission Line.—Part II." Ericsson Technics, no. 6 (1937).
(11) G. A. Campbell and R. M. Foster, " Fourier Integrals for Practical Applications," B. S. T. J. (Sept. 1931).
(12) Pipes, " The Operational Calculus."—Parts I., II., & III., Journal of Applied Physics (March, April, and May 1939).
(13) R. Courant und D. Hilbert, 'Methoden der Mathematischen Physick,' i. pp. 87–89 (Springer, 1924).

XXIII. *Relativity Theory of Electromagnetism.*

By Miss J. CATTERMOLE, Ph.D.*

[Received November 20, 1941.]

AMONG the attempts which have been made to include electromagnetic phenomena in a unified field theory, that of Kaluza (1) seems to the author to stand out as the most successful. Kaluza's theory has been developed by several authors, among whom are Klein (2), Rosenfeld and de Donder (3), Flint and Wilson (4), and Wilson and Cattermole (5). The advantage of Kaluza's theory over that of Weyl is considerable, for Weyl's theory merely houses electromagnetic phenomena, while the latter are inherent in Kaluza's theory. Just as gravitational phenomena are described by the geometry of the continuum of the general theory of relativity, so are electromagnetic phenomena described by the geometry of a continuum in a general theory of relativity extended to five dimensions.

In the ordinary general relativity theory the square of the line interval assumes the form

$$ds^2 = g_{ab}\, dx^a\, dx^b \quad \ldots\ldots\ldots\ldots \quad (1)$$

(repetition of Latin suffices implies summation over four indices).

In a region in which there is no field of force, the components g_{ab} of the Riemann curvature tensor are zero if $a \neq b$, and unity if $a = b$. Such a region is called " flat " space-time or Galilean space, and is characterized by ds^2 becoming the sum of squares

$$ds^2 = \sum_a dx^a = dx^2 + dy^2 + dz^2 + dw^2, \quad \ldots\ldots\ldots \quad (2)$$

where $w = ict$, $i = \sqrt{-1}\,|$.

If, however, a field of force is present in the region, the square of the line interval assumes the form given in (1). In the general theory of relativity extended to five dimensions, we describe a region of space as " Kaluzean " if the square of the line interval, $d\sigma$, assumes the form

$$d\sigma^2 = dx^2 + dy^2 + dz^2 + dw^2 + dx_5{}^2,$$

and such a region is one in which there is no field of force of any kind, neither gravitational nor electromagnetic.

In the special theory of relativity, a transformation from one set of co-ordinate axes to another, which takes place by a rotation in the x_1,

---

* Communicated by Allan Ferguson, M.A., D.Sc.

x_4 (or w) plane around the origin, serves to set a particle in motion with constant velocity along the x-axis, and the rotation is described by the equations of the Lorentz transformation. In the same way, in the Kaluzean special theory, a rotation in the x_5, x_4 plane around the origin serves to set the particle in motion with constant velocity along the fifth co-ordinate axis ; and this is equivalent to endowing the particle with an electric charge. The equations of the transformation, the exact counterpart of those of Lorentz, may be employed to work out the velocity of an electron (or positron) moving in the absence of an electromagnetic field, starting out from the velocity of a photon ; or to work out the second order wave equation for an electron (or positron) in the absence of an electromagnetic field from the second order wave equation for a photon.

It has been shown by Wilson and Cattermole (6) that expression of Maxwell's equations in five-dimensional form necessitates the addition of three equations, which turn out to be those which describe the spin of an electron, or positron, one equation which describes the conservation of electric charge, and three further equations of which the meaning is not yet fully understood, though they bear a clear relation to the hydrodynamical equations for the rotational motion of a fluid.

A further advantage of the five-dimensional theory is that it embraces the three conservation principles, viz., conservation of momentum, mass or energy, and electric charge, in one great principle of conservation of momentum in its wider sense. Of the five components of this more general momentum, conservation of the first three expresses conservation of momentum in the restricted sense ; that of the fourth conservation of mass or energy ; while that of the fifth expresses conservation of electric charge.

In the five-dimensional general theory of relativity the square of the line interval is given by

$$d\sigma^2 = \gamma_{\alpha\beta}\, dx^{\alpha}\, dx^{\beta},$$

in which the $\gamma_{\alpha\beta}$ are the components of the appropriate Riemann curvature tensor, and repetition of Greek suffices implies summation over five indices. In such a region a field of force exists, and this may be gravitational and electromagnetic in nature, solely gravitational, or, if gravitational forces may be neglected in comparison with electromagnetic forces, solely electromagnetic.

It is the purpose of this paper to show that the problem of working out the trajectory of a charged particle in an electric, or magnetic, or an electromagnetic field, can be solved using the general five-dimensional theory in the same way as the four-dimensional general theory can be used to work out the motion of a particle in a gravitational field.

When a charged particle moves in an electromagnetic field, its momentum p must be replaced by a more general momentum $\Pi = p + e\mathrm{K}$, where e is the charge on the particle. There are five components of Π, and the components K_x, K_y, K_z of K involved in Π_x, Π_y and Π_z are the components of the electromagnetic vector potential defined * by

$$\operatorname{curl} \mathrm{K} = \frac{\mathrm{H}}{a},$$

where H is the magnetic field intensity, and a is a constant depending upon the units chosen ($a=1$ for electromagnetic or electrostatic units, $a=c$ for mixed units—c being the velocity of light in empty space) and the permeability of the medium has been taken as unity.

The component of K involved in Π_4 is $\mathrm{K}_4 = \frac{i\mathrm{V}}{a}$, where V is the scalar potential. The fifth component $\Pi_5 = e\mathrm{K}_5$, where K_5 (regarded as a fifth component of K) is a constant having dimensions $[\mathrm{K}]^{-\frac{1}{2}}[\mathrm{L}]^{-\frac{1}{2}}[\mathrm{M}]^{-\frac{1}{2}}$.

In the general theory of relativity the momentum

$$p = m_0 g_{ab} \frac{dx^b}{d\tau} = m g_{ab} \frac{dx^b}{dt},$$

where m and m_0 are the mass and rest mass of the particle respectively, and $d\tau$ is the interval of proper time.

In the five-dimensional theory we put

$$\Pi_\alpha = \mu \gamma_{\alpha\beta} \frac{dx^\beta}{d\mathrm{T}} = m_0 \gamma_{\alpha\beta} \frac{dx^\beta}{d\tau} = m \gamma_{\alpha\beta} \frac{dx^\beta}{dt},$$

where μ and $d\mathrm{T}$ are the rest mass and interval of proper time appropriate to the continuum. We have

$$\Pi_5 = e\mathrm{K}_5 = m_0 \gamma_{5\beta} \frac{dx^\beta}{d\tau}.$$

$$\therefore \qquad e = \frac{m_0}{\mathrm{K}_5} \gamma_{5\beta} \frac{dx^\beta}{d\tau}.$$

Now $\quad \Pi_\alpha = m_0 \gamma_{\alpha\beta} \frac{dx^\beta}{d\tau} = m_0 g_{\alpha\beta} \frac{dx^\beta}{d\tau} + e\mathrm{K}_\alpha \quad$ (with $g_{5\alpha}=0$),

and substituting for e we have

$$\Pi_\alpha = m_0 g_{\alpha\beta} \frac{dx^\beta}{d\tau} + \frac{m_0 \mathrm{K}_\alpha}{\mathrm{K}_5} \gamma_{5\beta} \frac{dx^\beta}{d\tau}.$$

Equating coefficients of $m_0 \frac{dx^\beta}{d\tau}$, we have

$$\gamma_{\alpha\beta} = g_{\alpha\beta} + \frac{\mathrm{K}_\alpha}{\mathrm{K}_5} \gamma_{5\beta}.$$

---

* This is not a complete definition since imposing a gradient upon K would leave the curl unchanged.

Hence $$\gamma_{\beta 5}=\frac{K_\beta}{K_5}\gamma_{55}, \quad \text{since } g_{\alpha 5}=0,$$

or $$\gamma_{\alpha\beta}=g_{\alpha\beta}+\frac{K_\alpha K_\beta}{K_5^2}\gamma_{55}. \qquad (3)$$

We now proceed to investigate the motion of a charged particle under conditions in which gravitational forces are negligible compared with those of electric or magnetic character. Under these circumstances

$$g_{\alpha\beta}\begin{cases}=0 & \alpha\neq\beta,\\ =1 & \alpha=\beta,\end{cases}$$
$$g_{5\alpha}=0.$$

Motion of a Charged Particle under the Influence of a Magnetic Field.

We employ polar co-ordinates r, θ, ϕ, where

$$x=r\sin\theta\cos\phi,$$
$$y=r\sin\theta\sin\phi,$$
$$z=r\cos\theta,$$

and suppose the magnetic field to be in the direction of the z-axis so that $H_z=H$, $H_x=H_y=0$, and K is confined to planes parallel to the xy plane. The only component of K is $K_\varphi=\frac{1}{2}\frac{H}{a}r^2\sin^2\theta$ or $K_\phi=\frac{1}{2}Hr^2\sin^2\theta$ in electromagnetic units, so that the only surviving $\gamma_{\alpha\beta}$ are

$$\gamma_{11}=1;\quad \gamma_{22}=r^2;\quad \gamma_{33}=r^2\sin^2\theta\left(1+\frac{\gamma_{55}H^2r^2\sin^2\theta}{4K_5^2}\right);$$
$$\gamma_{35}=\gamma_{53}=\frac{Hr^2\sin^2\theta\,\gamma_{55}}{2K_5};\quad \gamma_{44}=1;\quad \gamma_{55}.$$

γ_{55} is unknown, but assumed constant for the purpose of this investigation.

Hence the square of the line interval reduces to

$$d\sigma^2=dr^2+r^2d\theta^2+r^2\sin^2\theta\left(1+\frac{\gamma_{55}H^2r^2\sin^2\theta}{4K_5^2}\right)d\phi^2$$
$$+2\left(\frac{Hr^2\sin^2\theta\gamma_{55}}{2K_5}\right)d\phi\,dx_5+dw^2+\gamma_{55}dx_5^2. \qquad (4)$$

The determinant $|\gamma|=\gamma_{55}r^4\sin^2\theta$ and the contra-variant $\gamma^{\alpha\beta}$

$$\left(\text{where } \gamma^{\alpha\beta}=\frac{\text{minor of } \gamma_{\alpha\beta}}{|\gamma|}\right)$$

are found to be

$$\gamma^{11}=1\,;\quad \gamma^{22}=\frac{1}{r^2};\quad \gamma^{33}=\frac{1}{r^2\sin^2\theta};\quad \gamma^{44}=1\,;\quad \gamma^{55}=\frac{1}{\gamma_{55}}+\frac{\mathrm{H}^2r^2\sin^2\theta}{4\mathrm{K}_5{}^2};$$

the remainder vanishing.

The geodesic equations are

$$\frac{d^2x_\alpha}{d\sigma^2}+\Gamma^\alpha_{\mu\nu}\frac{dx^\mu}{d\sigma}\cdot\frac{dx^\nu}{d\sigma}=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

in which the Christoffel symbols $\Gamma^\alpha_{\mu\nu}$ are given by

$$\Gamma^\alpha_{\mu\nu}=\tfrac{1}{2}\gamma^{\alpha\beta}\left(\frac{\partial\gamma_{\beta\mu}}{\partial x_\nu}+\frac{\partial\gamma_{\beta\nu}}{\partial x_\mu}-\frac{\partial\gamma_{\mu\nu}}{\partial x_\beta}\right). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Of the 125 Christoffel symbols, only 25 survive, the remainder being zero.

Considering first of all the second of the geodesic equations (5), namely

$$\frac{d^2\theta}{d\sigma^2}+\Gamma^2_{\mu\nu}\frac{dx^\mu}{d\sigma}\frac{dx^\nu}{d\sigma}=0,$$

this is

$$\frac{d^2\theta}{d\sigma^2}+2\Gamma^2_{12}\frac{dr}{d\sigma}\frac{d\theta}{d\sigma}+\Gamma^2_{33}\left(\frac{d\phi}{d\sigma}\right)^2+2\Gamma^2_{35}\frac{d\phi}{d\sigma}\frac{dx_5}{d\sigma}=0.$$

Calculating the Christoffel symbols from (6), we find

$$\frac{d^2\theta}{d\sigma^2}+\frac{2}{r}\frac{dr}{d\sigma}\frac{d\theta}{d\sigma}-\sin\theta\cos\theta\left(1+\frac{\gamma_{55}\mathrm{H}^2r^2\sin^2\theta}{2\mathrm{K}_5{}^2}\right)\left(\frac{d\phi}{d\sigma}\right)^2$$
$$-2\left(\mathrm{H}\frac{\gamma_{55}\sin\theta\cos\theta}{2\mathrm{K}_5}\right)\frac{d\phi}{d\sigma}\frac{dx_5}{d\sigma}=0.$$

Suppose the particle moves initially in the plane $\theta=\frac{\pi}{2}$, *i. e.*, the initial velocity is perpendicular to H and along the x-axis, then

$$\sin\theta=1,\quad \cos\theta=0,\quad \frac{d\theta}{d\sigma}=0,\quad \text{and} \therefore \frac{d^2\theta}{d\sigma^2}=0.$$

Hence $\frac{d\theta}{d\sigma}$ =constant, and ∴ remains zero always, θ remaining constant.

The five geodesic equations become

$$\frac{d^2r}{d\sigma^2}+\Gamma^1_{33}\left(\frac{d\phi}{d\sigma}\right)^2+2\Gamma^1_{35}\frac{d\phi}{d\sigma}\frac{dx_5}{d\sigma}=0. \quad . \quad . \quad . \quad . \quad . \quad (7)$$

$$\frac{d^2\theta}{d\sigma^2}=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

$$\frac{d^2\phi}{d\sigma^2}+2\Gamma^3_{13}\frac{dr}{d\sigma}\frac{d\phi}{d\sigma}+2\Gamma^3_{15}\frac{dr}{d\sigma}\frac{dx_5}{d\sigma}=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

$$\frac{d^2w}{d\sigma^2}=0. \qquad \qquad (10)$$

$$\frac{d^2x_5}{d\sigma^2}+2\Gamma^5_{13}\frac{dr}{d\sigma}\frac{d\phi}{d\sigma}+2\Gamma^5_{51}\frac{dr}{d\sigma}\frac{dx_5}{d\sigma}=0. \qquad \qquad (11)$$

From (10) we have, since $d\sigma^2=-c^2d\mathrm{T}^2$,

$$\frac{dw}{d\mathrm{T}}=\text{constant}=\beta \text{ say,}$$

and since

$$\mu\frac{dw}{d\mathrm{T}}=m\frac{dw}{dt}=mic=\mu\beta,$$

$$\therefore \qquad \beta=\frac{mic}{\mu} \quad \text{or} \quad \frac{dw}{d\mathrm{T}}=\frac{mic}{\mu}. \qquad \qquad (12)$$

Evaluating Γ^3_{13} and Γ^3_{15}, remembering that $\sin\theta=1$, (9) becomes

$$\frac{d^2\phi}{d\sigma^2}+2\left\{\frac{1}{r}+\frac{\gamma_{55}\mathrm{H}^2r}{4\mathrm{K}_5{}^2}\right\}\frac{dr}{d\sigma}\frac{d\phi}{d\sigma}+2\left\{\frac{\mathrm{H}\gamma_{55}}{2r\mathrm{K}_5}\right\}\frac{dr}{d\sigma}\frac{dx_5}{d\sigma}=0. \qquad (13)$$

Now, since

$$\Pi_5=e\mathrm{K}_5=\mu\gamma_{53}\frac{d\phi}{d\mathrm{T}}+\mu\gamma_{55}\frac{dx_5}{d\mathrm{T}};$$

$$\therefore \qquad \frac{d\phi}{d\mathrm{T}}=\frac{2e\mathrm{K}_5{}^2}{\mu\gamma_{55}\mathrm{H}r^2}-\frac{2\mathrm{K}_5}{\mathrm{H}r^2}\frac{dx_5}{d\mathrm{T}} \qquad \qquad (14)$$

and

$$\frac{d^2\phi}{d\mathrm{T}^2}=-\frac{4\mathrm{K}_5{}^2e}{\mu\gamma_{55}\mathrm{H}r^3}\frac{dr}{d\mathrm{T}}+\frac{4\mathrm{K}_5}{\mathrm{H}r^3}\frac{dr}{d\mathrm{T}}\frac{dx_5}{d\mathrm{T}}-\frac{2\mathrm{K}_5}{\mathrm{H}r^2}\frac{d^2x_5}{d\mathrm{T}^2}.$$

Substituting in (13) we find

$$\frac{d^2x_5}{d\mathrm{T}^2}=\frac{e\mathrm{H}^2r}{2\mu\mathrm{K}_5}\frac{dr}{d\mathrm{T}}.$$

$$\therefore \qquad \frac{dx_5}{d\mathrm{T}}=\frac{e\mathrm{H}^2r^2}{4\mu\mathrm{K}_5}+\text{constant.}$$

When $r=0$,

$$\mu\gamma_{55}\frac{dx_5}{d\mathrm{T}}=e\mathrm{K}_5,$$

hence

$$\text{constant}=\frac{e\mathrm{K}_5}{\mu\gamma_{55}},$$

and

$$\frac{dx_5}{d\mathrm{T}}=\frac{e\mathrm{H}^2r^2}{4\mu\mathrm{K}_5}+\frac{e\mathrm{K}_5}{\mu\gamma_{55}}, \qquad \qquad (15)$$

and, from (14),

$$\frac{d\phi}{d\mathrm{T}}=-\frac{e\mathrm{H}}{2\mu}. \qquad \qquad (16)$$

Dividing both sides of (4) by $d\sigma^2$, substituting for $\frac{dw}{d\mathrm{T}}$, $\frac{dx_5}{d\mathrm{T}}$ and $\frac{d\phi}{d\mathrm{T}}$ from (12), (15), and (16), and remembering that $\frac{d\theta}{d\mathrm{T}}=0$, we find

$$\left(\frac{dr}{d\mathrm{T}}\right)^2=1-\frac{e^2\mathrm{H}^2r}{4\mu^2}+\frac{m^2c^2}{\mu^2}-\frac{e^2\mathrm{K}_5{}^2}{\mu^2\gamma_{55}}.$$

When $r=0$,
$$\frac{dr}{dT}=\frac{m}{\mu}\frac{dr}{dt}=\frac{mv_0}{\mu},$$
where v_0 is the initial velocity of the charged particle.
$$\therefore \qquad \left(\frac{dr}{dT}\right)^2=\frac{m^2v_0{}^2}{\mu^2}-\frac{r^2e^2H^2}{4\mu^2},$$
and since from (16)
$$dT=-\frac{2\mu\, d\phi}{eH},$$
$$\left(\frac{dr}{d\phi}\right)^2=\frac{4v_0{}^2m^2}{e^2H^2}-r^2.$$
Putting $r=\frac{1}{u}$,
$$\left(\frac{du}{d\phi}\right)^2=\frac{4m^2v_0{}^2u^4}{e^2H^2}-u^2.$$

Differentiating with respect to ϕ. and cancelling out $2\frac{du}{d\phi}$, we obtain
$$\frac{d^2u}{d\phi^2}+u=\frac{8v_0{}^2m^2u^3}{e^2H^2}, \qquad . \; . \; . \; . \; . \; . \; . \; . \; . \qquad (17)$$
which is the equation of a circle of radius $\frac{mv_0}{eH}$, referred to the origin of projection as pole.

Hence the charged particle describes a circle of radius $\frac{mv_0}{eH}$ lying in the xz plane when acted upon by a magnetic field in the z direction.

Motion of a Charged Particle in an Electrostatic Field due to a Point Charge at the Origin.

Under these conditions, the only component of K is $K_4=\frac{iV}{c}$ using mixed units, where V is the scalar potential. As before, we employ polar co-ordinates, and at any point distant r from the origin, $V=\frac{B}{r}$ *, where B is a constant. Hence $K_4=\frac{iB}{cr}$. The co-variant $\gamma_{\alpha\beta}$, which do not vanish, are
$$\gamma_{11}=1\,; \quad \gamma_{22}=r^2\,; \quad \gamma_{33}=r^2\sin^2\theta\,;$$
$$\gamma_{44}=1-\frac{B^2\gamma_{55}}{c^2K_5{}^2r^2}; \quad \gamma_{54}=\gamma_{45}=\frac{iB\gamma_{55}}{cK_5r}; \quad \gamma_{55}.$$

The determinant $|\gamma|=\gamma_{55}r^4\sin^2\theta$, and the non-vanishing contra-variant $\gamma^{\alpha\beta}$, are

---

* *Footnote added in proof.*—It may be necessary to add a small correcting term to B/r in subsequent investigations, but $V=B/r$ is assumed as a good approximation for present purposes.

$$\gamma^{11}=1\,;\quad \gamma^{22}=\frac{1}{r^2};\quad \gamma^{33}=\frac{1}{r^2\sin^2\theta};\quad \gamma^{44}=1\,;$$

$$\gamma^{45}=\gamma^{54}=-\frac{iB}{crK_5};\quad \gamma^{55}=\frac{1}{\gamma_{55}}-\frac{B^2}{c^2K_5{}^2r^2}.$$

The non-vanishing Christoffel symbols are twenty in number, namely

$$\Gamma^1_{22}=-r\,;\quad \Gamma^1_{33}=-r^2\sin^2\theta\,;\quad \Gamma^1_{44}=-\frac{B^2\gamma_{55}}{c^2K_5{}^2r^3};\quad \Gamma^1_{45}=\Gamma^1_{54}=\frac{iB\gamma_{55}}{2cK_5r^2};$$

$$\Gamma^2_{12}=\Gamma^2_{21}=\frac{1}{r};\quad \Gamma^2_{33}=-\sin\theta\cos\theta\,;\quad \Gamma^3_{13}=\Gamma^3_{31}=\frac{1}{r};\quad \Gamma_{23}=\Gamma^3_{32}=\cot\theta;$$

$$\Gamma^4_{14}=\Gamma^4_{41}=\frac{B^2\gamma_{55}}{2K_5{}^2r^3c^2};\quad \Gamma^4_{15}=\Gamma^4_{51}=-\frac{iB\gamma_{55}}{2K_5cr^2};$$

$$\Gamma^5_{14}=\Gamma^5_{41}=-\frac{iB^3\gamma_{55}}{2K_5{}^3c^3r^4}-\frac{iB}{2K_5cr^2};\quad \Gamma^5_{15}=\Gamma^5_{51}=-\frac{B^2\gamma_{55}}{2c^2K_5{}^2r^3}.$$

Of the geodesic equations, consider the second,

$$\frac{d^2\theta}{d\sigma^2}+\frac{2}{r}\frac{dr}{d\sigma}\frac{d\phi}{d\sigma}+2\cot\theta\frac{d\theta}{d\sigma}\frac{d\phi}{d\sigma}=0.$$

If the particle moves initially in the plane $\theta=\frac{\pi}{2}$ it is easy to see that $\frac{d^2\theta}{d\sigma^2}$ and $\frac{d\theta}{d\sigma}$ are always zero, and θ remains $\frac{\pi}{2}$.

Hence the geodesic equations are

$$\frac{d^2r}{d\sigma^2}-r\left(\frac{d\phi}{d\sigma}\right)^2-\frac{B^2\gamma_{55}}{K_5{}^2c^2r^3}\left(\frac{dw}{d\sigma}\right)^2+\frac{iB\gamma_{55}}{K_5cr^2}\frac{dw}{d\sigma}\frac{dx_5}{d\sigma}=0.\quad . \quad . \quad (18)$$

$$\frac{d^2\theta}{d\sigma^2}=0.\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

$$\frac{d^2\phi}{d\sigma^2}+\frac{2}{r}\frac{dr}{d\sigma}\frac{d\phi}{d\sigma}=0.\quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

$$\frac{d^2w}{d\sigma^2}+\frac{B^2\gamma_{55}}{c^2K_5{}^2r^3}\frac{dr}{d\sigma}\frac{dw}{d\sigma}-\frac{iB\gamma_{55}}{CK_5r^2}\frac{dr}{d\sigma}\frac{dx_5}{d\sigma}=0.\quad . \quad . \quad . \quad (21)$$

$$\frac{d^2x_5}{d\sigma^2}-\left(\frac{iB^3\gamma_{55}}{c^3K_5{}^3r^4}+\frac{iB}{cK_5r^2}\right)\frac{dr}{d\sigma}\frac{dw}{d\sigma}-\frac{B^2\gamma_{55}}{c^2K_5{}^2r^3}\frac{dr}{d\sigma}\frac{dx_5}{d\sigma}=0.\quad . \quad (22)$$

Integrating (20) we find $\quad \frac{d\phi}{dT}=\frac{\alpha}{r^2},$

where α is a constant, and since

$$\mu\frac{d\phi}{dT}=m\frac{d\phi}{dt}.$$

we have $\quad \alpha=\frac{\Omega}{\mu},\quad$ where $\Omega=mr^2\frac{d\phi}{dt}$.

the angular momentum of the particle;

$$\therefore \qquad \frac{d\phi}{dT}=\frac{\Omega}{\mu r^2}. \qquad (23)$$

Now $$\mu\gamma_{54}\frac{dw}{dT}+\mu\gamma_{55}\frac{dx_5}{dT}=eK_5.$$

$$\therefore \qquad \frac{dx_5}{dT}=\frac{eK_5}{\mu\gamma_{55}}-\frac{iB}{cK_5 r}\frac{dw}{dT}.$$

Substituting in (21), we have

$$\frac{d^2w}{dT^2}-\frac{iBe}{\mu cr^2}\frac{dr}{dT}=0.$$

Hence $$\frac{dw}{dT}=-\frac{iBe}{\mu cr}+\lambda, \qquad (24)$$

where λ is a constant of integration.

Also $$\frac{dx_5}{dT}=\frac{eK_5}{\mu\gamma_{55}}-\frac{iB\lambda}{cK_5 r}+\frac{B^2e}{\mu c^2K_5 r^2}. \qquad (25)$$

The square of the line interval assumes the form

$$d\sigma^2=dr^2+r^2d\phi^2+\frac{2iB\gamma_{55}}{cK_5 r}dw\,dx_5+\left(1-\frac{B^2\gamma_{55}}{c^2K_5^{\,2}r^2}\right)dw^2+\gamma_{55}dx_5^{\,2},$$

and dividing by dT^2, substituting for $\frac{d\phi}{dT}, \frac{dw}{dT}, \frac{dx_5}{dT}$ from (23), (24) and (25), remembering that

$$dT=\frac{\mu r^2 d\phi}{\Omega},$$

we obtain $$-c^2=\frac{\Omega^2}{\mu^2r^4}+\frac{\Omega^2}{\mu^2r^2}+\lambda^2-\frac{B^2e^2}{\mu^2c^2r^2}-\frac{2\lambda iBe}{\mu cr}+\frac{e^2K_5^{\,2}}{\mu^2\gamma_{55}}.$$

Putting $r=\frac{1}{u}$, and differentiating with respect to ϕ, we obtain

$$\frac{d^2u}{d\phi^2}+u\left(1-\frac{B^2e^2}{\Omega^2}\right)=\frac{\lambda iBe\mu}{\Omega^2}. \qquad (26)$$

The general solution is

$$u=\frac{\lambda iBe\mu}{\Omega^2-B^2e^2}+R\cos\left\{\sqrt{1-\frac{B^2e^2}{\Omega^2}}\,\phi-\eta\right\}, \qquad (27)$$

where R and η are constants of integration.

This is the equation of a conic processing in the plane of motion in the same direction as that in which the particle is moving, since $\sqrt{1-\frac{B^2e^2}{\Omega^2}}$ is less than unity.

Motion of a Charged Particle under the Simultaneous Action of a Central Electrostatic Field and a Magnetic Field.

As in the case of motion in a magnetic field only, previously considered, we suppose the magnetic field to lie in the z direction, and use polar co-ordinates as before.

We now have two components of K, namely

$$\left.\begin{aligned} K_\phi &= \frac{1}{2}\frac{H}{c}r^2\sin^2\theta, \\ K_4 &= \frac{iB}{cr} \end{aligned}\right\} \text{using mixed units } (a=c).$$

The non-vanishing co-variant $\gamma_{\alpha\beta}$ are :

$$\gamma_{11}=1\,;\quad \gamma_{22}=r^2\,;\quad \gamma_{33}=r^2\sin^2\theta\left(1+\frac{\gamma_{55}H^2r^2\sin^2\theta}{4K_5{}^2c^2}\right);$$

$$\gamma_{34}=\gamma_{43}=\frac{iBHr\sin^2\theta\gamma_{55}}{2K_5{}^2c^2};\quad \gamma_{35}=\gamma_{53}=\frac{Hr^2\sin^2\theta\gamma_{55}}{2K_5c};$$

$$\gamma_{44}=1-\frac{B^2\gamma_{55}}{K_5{}^2c^2r^2};\quad \gamma_{45}=\gamma_{54}=\frac{iB\gamma_{55}}{cK_5r};\quad \gamma_{55}.$$

The determinant $|\gamma|=\gamma_{55}r^4\sin^2\theta$, and the non-vanishing contra-variant $\gamma^{\alpha\beta}$ are

$$\gamma^{11}=1\,;\quad \gamma^{22}=\frac{1}{r^2};\quad \gamma^{33}=\frac{1}{r^2\sin^2\theta};\quad \gamma^{35}=\gamma^{53}=-\frac{H}{2K_5c};$$

$$\gamma^{44}=1\,;\quad \gamma^{45}=\gamma^{54}=-\frac{iB}{crK_5};\quad \gamma^{55}=\frac{1}{\gamma_{55}}-\frac{B^2}{K_5{}^2r^2c^2}+\frac{Hr^2\sin^2\theta}{4K_5{}^2c}.$$

There are now 46 non-vanishing Christoffel symbols, but for present purposes it will not be necessary to calculate them all. Those which do not vanish, and which are concerned in the second geodesic equation

$$\frac{d^2\theta}{d\sigma^2}+\Gamma^2_{\alpha\beta}\frac{dx^\alpha}{d\sigma}\frac{dx^\beta}{d\sigma}=0$$

are $$\Gamma^2_{21}=\Gamma^2_{12}=\frac{1}{r};\quad \Gamma^2_{33}=-\sin\theta\cos\theta\left(1+\frac{\gamma_{55}H^2r^2\sin^2\theta}{2K_5{}^2c^2}\right);$$

and $$\Gamma^2_{35}=\Gamma^2_{53}=-\frac{H\gamma_{55}\sin\theta\cos\theta}{2K_5c},$$

so that if the particle moves initially in the plane $\theta=\frac{\pi}{2}$ then $\frac{d^2\theta}{d\sigma^2}=0$. $\frac{d\theta}{d\sigma}$ is constant $=0$ always, and θ is constant as before.

The third and fourth geodesic equations after evaluation of the non-vanishing Christoffel symbols reduce to

$$\frac{d^2\phi}{dT^2}+\left(\frac{2}{r}+\frac{\gamma_{55}H^2r}{2K_5{}^2c^2}\right)\frac{dr}{dT}\frac{d\phi}{dT}+\frac{H\gamma_{55}}{crK_5}\frac{dr}{dT}\frac{dx_5}{dT}+\frac{iBH\gamma_{55}}{K_5{}^2r^2c^2}\frac{dr}{dT}\frac{dw}{dT}=0. \qquad (28)$$

$$\frac{d^2w}{dT^2}+\frac{B^2\gamma_{55}}{K_5{}^2c^2r^3}\frac{dr}{dT}\frac{dw}{dT}-\frac{iB\gamma_{55}}{K_5cr^2}\frac{dr}{dT}\frac{dx_5}{dT}-\frac{iBH\gamma_{55}}{2c^2K_5{}^2}\frac{d\phi}{dT}\frac{dr}{dT}=0. \quad (29)$$

Now $\Pi_5=eK_5=\mu\gamma_{5\alpha}\frac{dx^\alpha}{dT}=\mu\gamma_{53}\frac{d\phi}{dT}+\mu\gamma_{54}\frac{dw}{dT}+\mu\gamma_{55}\frac{dx_5}{dT}.$

$$\therefore \quad \frac{dx_5}{dT}=\frac{eK_5}{\mu\gamma_{55}}-\frac{Hr^2}{2K_5c}\frac{d\phi}{dT}-\frac{iB}{cK_5r}\frac{dw}{dT}. \quad . \quad . \quad . \quad . \quad . \quad (30)$$

From (28) we find

$$\frac{d^2\phi}{dT^2}+\frac{2}{r}\frac{dr}{dT}\frac{d\phi}{dT}+\frac{eH}{\mu rc}\frac{dr}{dT}=0.$$

$$\therefore \quad \frac{d}{dT}\left(r^2\frac{d\phi}{dT}\right)+\frac{eHr}{\mu c}\frac{dr}{dT}=0.$$

$$\therefore \quad \frac{d\phi}{dT}=-\frac{eH}{2\mu c}+\frac{\alpha}{r^2},$$

α being a constant of integration.

Hence the angular momentum

$$\frac{d\phi}{dt}=\frac{\mu}{m}\frac{d\phi}{dT} \quad \text{or} \quad \frac{d\phi}{dt}=-\frac{eH}{2mc}+\frac{\alpha\mu}{mr^2}.$$

In the absence of the magnetic field, H=0, and the angular velocity $=\frac{\alpha\mu}{mr^2}$, so that the presence of the magnetic field necessitates the addition of the angular velocity $\left(-\frac{eH}{2mc}\right)$ to the angular velocity of the charged particle in the absence of a magnetic field. $\frac{eH}{2mc}$ is the Larmor precessional velocity.

In the light of the results of these investigations, taken together with its successes mentioned earlier in this paper, there seems to be no doubt that the Kaluza extension of the general theory of relativity to five dimensions renders a more complete account than any other unified field theory previously presented.

References.

(1) Kaluza, "Zum Unitätsproblem der Physik," *Preuss. Akad. d. Wiss. Berlin*, liv. p. 966 (1921).

(2) Klein, *Zeits. für Physik*, xxxvii. p. 895 (1926).

(3) Rosenfeld and de Donder, Series of papers in *Ac. Roy. Belg. Bull. Class. Sci.* 5th series, 1927–1930.

(4) Flint and Wilson, Phys. Soc. l. p. 340 (1938).

(5) Wilson and Cattermole, Phil. Mag. ser. 7, xxvii. p. 84 (1939).

(6) *Loc. cit.*

XXIV. *The Formula of the Selenium Barrier-Layer Photo-Cell.*

By R. A. HOUSTOUN, D.Sc., University of Glasgow *.

[Received December 29, 1941.]

IN using the selenium barrier-layer cell it is connected to a microammeter and resistance, the combined resistance of which is R, and exposed to an illumination of L foot-candles. The current I, recorded by the microammeter, is then a function of L and R. Curiously enough, no one seems to have investigated the form of this function. The present paper is an attempt to fill the gap. It is known, of course, that when R is small, I is simply proportional to L, and the makers of the cells state this constant of proportionality, but there is no comparison on record of the performances of different modern cells from an independent source and no account of the effect of varying R, except in the book of data supplied with the Photronic cell.

Fig. 1.

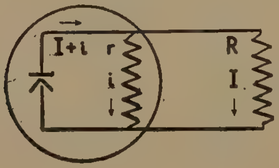
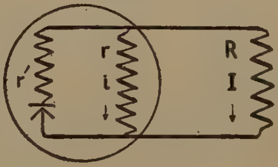


Fig. 2.



According to the theory given in this book, the action of the cell may be represented in fig. 1. R is the resistance of the external circuit, including the microammeter, and I the current which flows through R; r is the internal resistance of the cell and i a current which discharges back through r. The total current $I+i$ is proportional to the incident light and is independent of R. If R is very small, i may be disregarded. Hence, by measuring I for different values of L and R, r may be determined as a function of L and R, and its variation with these two quantities is shown by a series of curves.

I have found it better to take a different view of the action of the cell, one that is represented in fig. 2. The e.m.f. is supposed to be constant and of value E, but there is a variable resistance r' which is inversely proportional to the intensity of illumination,

---

* Communicated by the Author.

$$r' = \frac{k}{L},$$

where k is a constant and r' and E are measured respectively in ohms and volts. r, i, R and I have the same meanings as before.

$$ri = RI$$

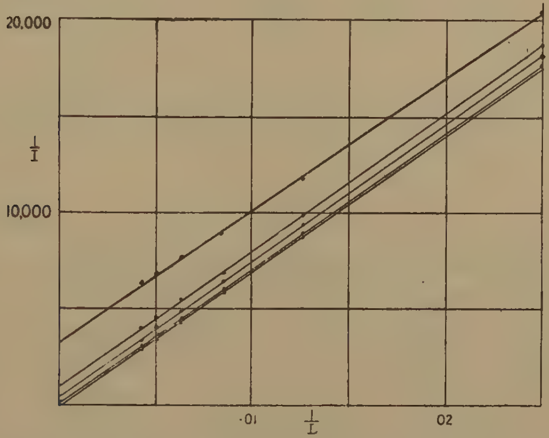
and

$$E = r'(I+i) + RI.$$

These equations give

$$\frac{1}{I} = \frac{k}{E}\left(1+\frac{R}{r}\right)\frac{1}{L} + \frac{R}{E},$$

Fig. 3.



Taking the lines in order from the top of the diagram,
R is 1000, 500, 300, 100 and 3 ohms.

so that 1/I should give a straight line when graphed against 1/L, provided that R and r are constant.

If we take the view represented in fig. 1 and write

$$I + i = \frac{L}{k'},$$

we find that

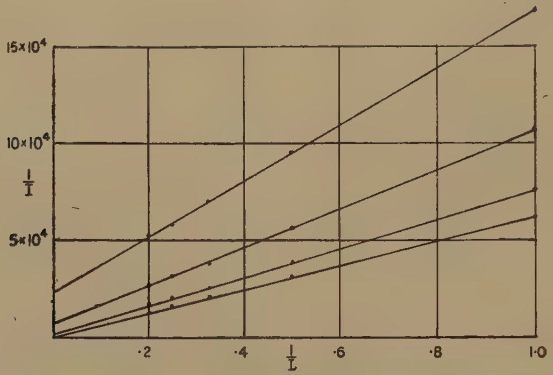
$$\frac{1}{I} = k'\left(1+\frac{R}{r}\right)\frac{1}{L},$$

also a straight line under the same condition, but a straight line which goes through the origin.

In fig. 1 of the Photronic book there are curves giving I in terms of L for seven different values of R, for values of L up to 250 foot-candles, and in fig. 2 there are curves for twelve other values of R for the range 0 to 6 foot-candles. I have re-drawn five of the first set and four of the second set as 1/I in terms of 1/L in figs. 3 and 4 of the present paper. I have omitted the other curves in order not to crowd unduly the diagram. There is no doubt whatever that the graphs are straight lines and that they do not go through the origin.

The intercept on the 1/I axis is R/E. If it is measured, E can be calculated. The results below were obtained in this manner.

Fig. 4.



Taking the lines in order from the top of the diagram, R is 24,000, 10,000. 4000 and 0 ohms.

	R.	E.
Fig. 3	1,000 ohms 500 300 100	0·31 volt 0·50 0·60 0·50
Fig. 4	24,000 ohms 10,000 4,000	1·09 volts 1·43 2·00

They vary widely. It should be noted, however, that the largest is about six times the smallest, whereas the values of R on which they are based vary 240 times, also that the values given are in each case the first that came. It would be possible to re-draw the lines so as to give a much better agreement. As a mean 0·60 volt was selected.

When R is zero, the gradient is $k/E = 7{\cdot}04 \times 10^5$. Hence $k = 4{\cdot}224 \times 10^5$, and over the range investigated r' varies from infinity to 1700 ohms.

The ratio of the gradient to the gradient at $R=0$ is $1+R/r$. The only conclusion to be drawn from fig. 3 is that r is much greater than 1000 ohms. From fig. 4 we obtain

R.	$1+\frac{R}{r}$.	r.
4,000 ohms	1·20	20,000 ohms
10,000	1·60	16,700
24,000	2·38	17,400

but the observations are apparently not very accurate, and all that we can deduce is that r is fairly constant over the range. The variation in r recorded at the end of the present paper, about which there is no doubt, occurs at lower values of R.

In order to investigate the variation of r with R more thoroughly, 1/I was graphed as a function of R, keeping L constant, for selected values of L, using the data given in figs. 1 and 2 of the Photronic book. Rearranging the equation as a function of R

$$\frac{1}{I}=\frac{k}{EL}+\frac{R}{E}\left(1+\frac{k}{rL}\right),$$

and it was expected that this would satisfy the data, if r were written $f(R)$. A value of $f(R)$ was found that satisfied the observations up to a point, but there were deviations, and I had doubts as to whether it was permissible to treat the data in this way, *i. e.*, if the curves were intended to represent I as a function of L for various values of R, it might not be fair to expect them to represent it as a function of R to a high degree of accuracy. So I resolved to make observations of my own.

I had at my disposal five cells, the particulars of which are as follows :—

A. A Weston Photronic which has been in use eight years.

B. A Weston Photronic obtained a year ago. The sensitive area of these cells is circular, 37 mm. in diameter, and it is protected by glass.

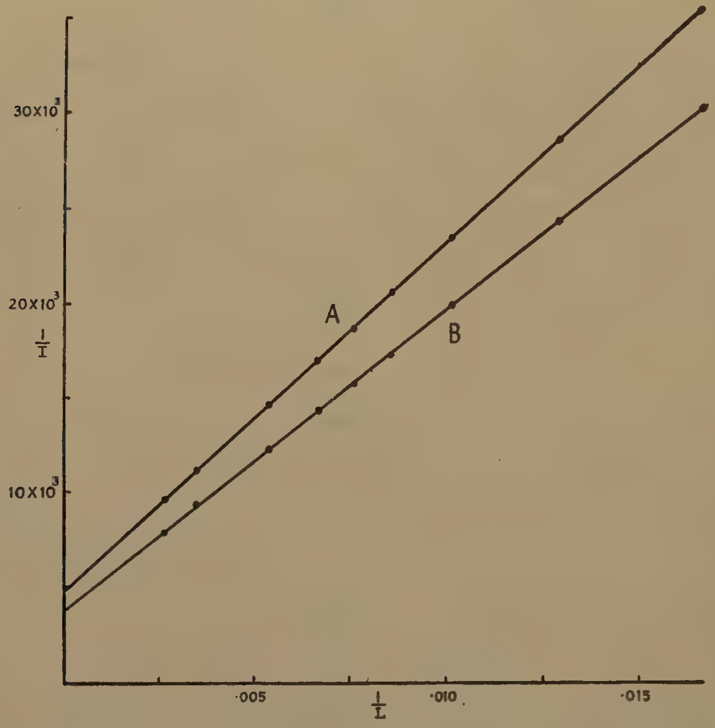
C. A cell supplied by Renny and Wender, Ltd. Area circular, 38 mm. in diameter. Made in England since the war. No glass cover.

D. Cell supplied by Evans Electro-selenium, Ltd. Rectangular, 30×13 mm.^{2}. Made in England since the war. No glass cover.

E. Cell of unknown origin, probably German. Rectangular, 30×15 mm.^{2}. No glass cover.

These were exposed on a photometer bench to a 100 watt, clear, gas-filled lamp and the current measured at various distances, steps being taken to keep the voltage on the lamp constant. Its candle-power was found to be 140 international candles by comparison with a Hefner lamp. The results are given in the following table (Table I.). The first column gives the illumination on the sensitive area in foot-candles. Columns 2, 4, 6, 8, 10 give the current recorded by cells A, B, C, D and E respectively, when there was nothing in the circuit but the microammeter, which had a resistance of 39·4 ohms. Columns 3, 5, 7, 9 and 11 give the current recorded by the same cells when there was a resistance in the circuit. The value of this resistance is stated at the top of the appropriate column. The currents are given in microampères. In all cases the full area of the cell was exposed to the illumination and the latter was, of course, uniform.

Fig. 5.



When the currents are plotted against the foot-candles for the case of no external resistance, the curve is a straight line except in the case of cell C. Here, and in the five cases where there is an external resistance, the current falls away from the straight line at high illuminations. But when the reciprocal of the current is plotted against the reciprocal of the

TABLE I.—Variation of I with L.

L.	A.		B.		C.		D.		E.	
		1000 ohms		1000 ohms		400 ohms		2000 ohms		1400 ohms
2·14	1·95	1·50	2·02	1·90	1·15	1·0	3·0	3·0	2·9	2·5
6·07	4·37	3·25	5·02	3·70	2·95	2·1	7·3	6·5	7·1	6·0
14·0	9·85	7·22	11·15	8·67	6·50	4·5	16·2	14·7	15·6	12·7
22·3	16·10	11·45	17·6	13·70	10·35	6·9	25·8	21·6	25·1	19·7
40·9	29·15	20·20	31·5	24·87	18·8	12·2	46·7	36·0	45·4	32·7
60·4	43·20	28·30	47·3	33·33	26·8	17·5	70·1	49·1	67·1	44·7
75·9	53·57	34·65	59·4	40·82	32·7	21·1	88·6	58·3	84·1	52·7
98·2	68·7	42·45	76·8	49·60	40·9	25·2	113·9	68·3	109·1	61·4
116·6	82·0	48·45	92·2	56·95	47·0	28·5	. .	76·0	. .	68·1
132·0	92·9	53·30	103·6	62·60	52·1	31·1	. .	81·2	. .	73·3
150·5	105·7	58·15	119·0	68·47	57·6	34·0	. .	87·1	. .	77·7
186·6	. .	67·45	. .	79·57	67·7	39·5	. .	96·0	. .	86·1
284·1	. .	88·10	. .	102·9	92·9	51·8	. .	114·2	. .	102·7
384·3	. .	103·0	. .	119·0	112·9	61·2	. .	. .	. .	113·3

illumination, the curve is in every case a straight line. This is shown in figs. 5, 6 and 7; fig. 5 is for cells A and B, fig. 6 is for the two cases of cell C, and fig. 7 is for cells D and E.

Fig. 6.

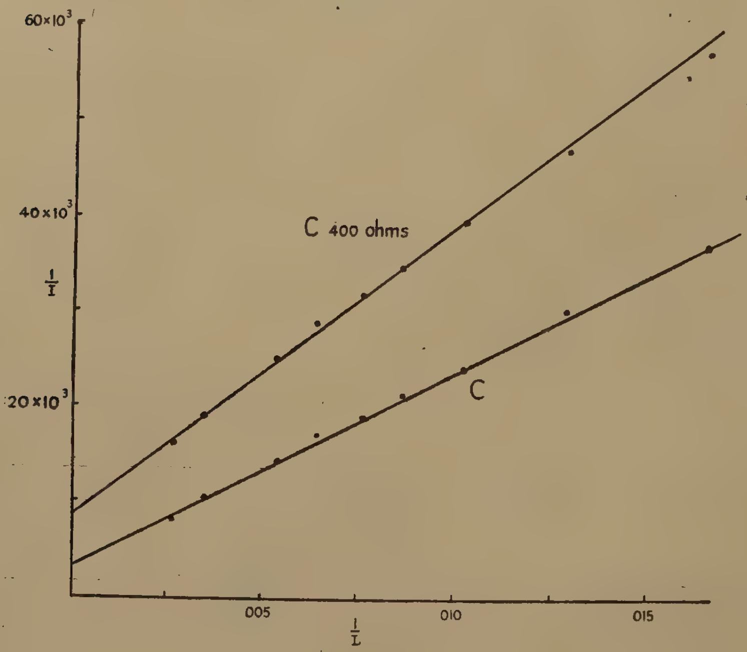
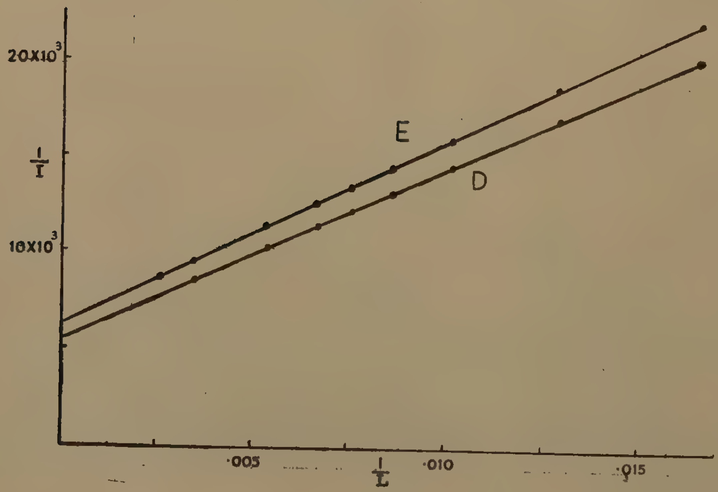


Fig. 7.



From the intercept on the axis R/E, the gradient for no external resistance k/E, and the ratio of the gradients $1+\mathrm{R}/r$, we derive some interesting data. These are given in the following table :—

Cell.	Intercept.	Gradient.	E.	r.	r'.
A	$4{\cdot}9\times10^{3}$	$12{\cdot}25\times10^{5}$ $18{\cdot}5\times10^{5}$	0·20 volt	3,300 ohms	7×10^{6} ohms
B	$4{\cdot}3\times10^{3}$	$12{\cdot}7\times10^{5}$ $15{\cdot}5\times10^{5}$	0·23 ,,	4,500 ,,	$5{\cdot}4\times10^{6}$,,
C	$3{\cdot}3\times10^{3}$ $8{\cdot}8\times10^{3}$	$20{\cdot}6\times10^{5}$ $30{\cdot}2\times10^{5}$	0·73 ,,	618 ,,	$10{\cdot}8\times10^{6}$,,
D	$5{\cdot}5\times10^{3}$	$8{\cdot}6\times10^{5}$ $9{\cdot}0\times10^{5}$	0·36 ,,	43,000 ,,	$2{\cdot}3\times10^{6}$,,
E	$6{\cdot}3\times10^{3}$	$9{\cdot}2\times10^{5}$ $9{\cdot}7\times10^{5}$	0·22 ,,	26,000 ,,	$4{\cdot}1\times10^{6}$,,

The value of r', of course, varies inversely as the illumination ; the value given above is for one foot-candle, *i. e.*, it is equal to k. The behaviour of the cell C can be explained only on the assumption that it has an "external resistance" R of about 240 ohms inside the cell, probably in the translucent metal film which covers the selenium. It is on this assumption that the values for this cell are calculated. The wide variation in the values of r should be noted. C is emphatically a low resistance cell and D and E are high resistance cells.

The values for E also vary widely, but in this case a large part of the variation is due to error in observation, how much it would require a separate investigation to determine. The cells, with the exception of D, are subject to fatigue at high illuminations, as I have shown in a former paper *, and it is the values of I at high illuminations which are important for determining the intercept. In order to see what accuracy is possible, some determinations of the intercept were made for cell D for red and green light. The red was ordinary signal red and the green was Chance's tricolour green. The following figures multiplied by 10^{3} give the results :—

Green	6·0,	6·1,	6·5,	6·7,	6·5,	6·4,	6·6,	6·3,	6·2,
Red	6·1,	8·4,	6·4,	6·4,	8·5,	9·8,	9·0,	7·5,	6·0.

Each results rests on two measurements of current. The wide variation for red as contrasted with green is due to the currents for red being much smaller, and probably also to the fact that the red glass filter did not absorb equally all over and that different portions of it may have been in front of the cell at different times. It may be possible to eliminate the disturbing factors and establish a variation of E with the wave-length.

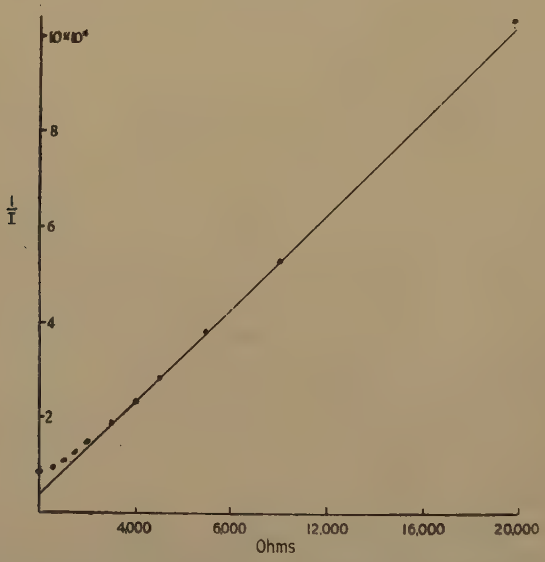
---

* Phil. Mag. xxxi. p. 498 (1941).

A lower limit to the value of E can be obtained by connecting the cell to a high resistance voltmeter. According to a leaflet issued in connexion with the Tungsram cell, when tested in this way it gives 0·4 volt for an illumination of 1000 lux, *i. e.*, about 91 foot-candles.

After the variation of I with L was studied, I proceeded to measure its variation with R, and the values of I in Table II. were obtained by keeping the photo-cell at a fixed position on the bench and varying the resistance in the circuit. They are expressed in microampères. Each photo-cell was used in two positions on the bench and the number at the top of the column gives the illumination for the position in question in foot-candles. If we graph 1/I as a function of R and assume that r is constant, the curve should be a straight line. As a matter of fact, when this is done, we obtain straight lines in the case of the two sets of readings recorded for cell C. For this cell r is constant. For the other cells there is a deviation from the straight line at low values of R. This deviation is greatest for cell D, and in order to show its nature the first set of results for this cell is graphed in fig. 8. The ordinates are expressed in amp.$^{-1}$ and the abscissæ in ohms. The observations can be represented by

Fig. 8.



$$\frac{1}{I} = \frac{k}{EL} + \frac{R}{E}\left(1 + \frac{k}{rL}\right) + ae^{-R/b},$$

TABLE II.—Variation of I with R.

Res.	A.		B.		C.		D.		E.	
	124·0.	72·3.	116·6.	69·0.	345·5.	150·5.	98·15.	49·2.	98·15.	49·2.
0 ohms	112·7	63·0	120·8	68·7	106·1	54·0	116·5	57·9	120·9	61·0
50 ,, ...	..	..	..	..	89·6	..	..	..	..	..
100 ,, ...	..	..	..	..	82·2	45·4	..	..	..	..
200 ,, ...	..	..	..	..	69·2	39·5	..	..	..	..
500 ,, ...	84·2	50·0	93·3	56·8	50·0	28·2	105·4	54·7	100·0	53·7
1,000 ,, ...	63·0	40·0	69·7	45·8	33·1	19·6	92·2	51·0	78·5	46·1
1,500 ,, ...	49·2	..	54·5	..	24·8	14·75	77·7	..	62·7	..
2,000 ,, ...	40·3	27·3	44·0	31·0	20·0	..	68·2	42·5	51·4	34·1
3,000 ,, ...	28·9	..	31·5	..	14·3	8·5	52·3	..	37·5	..
4,000 ,, ...	22·8	16·4	24·8	18·6	11·0	..	41·9	29·7	29·2	29·2
5,000 ,, ...	18·9	..	20·4	..	9·00	5·4	34·8	..	24·0	..
7,000 ,, ...	13·8	10·0	14·7	11·4	6·60	..	26·1	19·9	17·7	21·5
10,000 ,, ...	9·75	7·15	10·45	8·05	4·70	3·0	18·8	14·6	12·6	9·75
20,000 ,, ...	5·00	3·90	5·35	4·25	..	..	9·7	7·7	6·5	5·00

where the last term on the right is a purely empirical one, introduced to represent the deviation, and a and b are constants.

The two quantities k and r, which might be regarded as variables, clearly do not vary with L, as is shown by the observations recorded in Table I. The fact that it is the low resistance cell C which has a constant value of r makes it plausible that the deviation is connected with the value of r. We can therefore regard the additional term as due to a variation in r, but it does not seem worth while to re-write the formula and make it appear as such.

The variation is a decrease in the value of r when R is small, *i. e.*, when the return current is almost entirely an external one. Why r should decrease under these conditions it is impossible to say, until we have more information at our disposal.

Note added on correcting the proofs.—The formula verified above was obtained by trial, but it may be justified theoretically in the following manner:—

The cell forms a condenser which is discharging externally through a resistance R and internally through a resistance r. The rate of discharge

$$I+i=I\left(1+\frac{R}{r}\right).$$

The potential difference between the plates of the condenser is IR. Let the number of electrons liberated per second between the plates be aL, where L is, as before, the intensity of illumination and a is a constant. These will have varying degrees of energy; we assume that their law of distribution is

$$ce^{-cV}dV,$$

where V is the potential difference the electron can get over. V varies from 0 to E where E is the e.m.f. of the cell. The fraction available to charge the condenser is

$$\int_{IR}^{E} ce^{-cV}dV \text{ divided by } \int_{0}^{E} ce^{-cV}dV$$

or

$$\frac{e^{-cIR}-e^{-cE}}{1-e^{-cE}}.$$

The rate of charge is thus multiplied by aL. It becomes, on taking the first two terms of the expansions,

$$\frac{L}{k}(E-IR),$$

where k is a constant given by

$$k=(1-e^{-cE})/ac.$$

On equating the rate of charge to the rate of discharge

$$I\left(1+\frac{R}{r}\right)=\frac{L}{k}(E-IR),$$

which is the formula already obtained. If we take the next terms in the expansions we obtain a deviation in the right direction when R is low, but not of the right magnitude, but it is obvious that this could be put right by adjusting the law of distribution of energy of the electrons.

XXV. *Notices respecting New Books.*

Fourier Series and Boundary Value Problems. By Ruel V. Churchill. [Pp. x+206.] (New York and London: McGraw-Hill, 1941. Price 17*s.* 6*d.*)

Boundary value problems are of such frequent occurrence in physics that a knowledge of the mathematical techniques for dealing with them would be considered useful by the most practically-minded physicist. At the same time the study of these problems in their broader aspects leads to fields of investigation which can rouse the interest of the purest mathematician. Here, as elsewhere, the physicist of average mathematical attainment has difficulty in finding a treatment suited to his requirements. The available books on applicable mathematics tend to be either so elementary as to be almost useless in connexion with actual physical problems, or so advanced as to be almost beyond the powers of any but the professional mathematician. The present book provides a treatment of the particular field which is of just the scope and standard which is most needed.

After introducing the related problems of the expansion of a function in infinite series of prescribed types, and of the boundary value problems in partial differential equations, the author discusses the classical partial differential equations of physics as they arise in connexion with gravitational and electric potential, with heat flow, and with mechanical vibration. The properties of orthogonal sets of functions are then considered. After this preparation, a systematic presentation is given of Fourier series and integrals, and of their application in the solution of boundary value problems. A chapter is devoted to theorems connected with the uniqueness of solutions. The last two chapters deal with Bessel functions and Legendre polynomials and their applications.

It should be stated that the book is not primarily addressed to physicists. The physical problems considered are of that idealized type in which the physics is adapted to the mathematics, the interest lying in the mathematical methods rather than in the physical results. The power of these methods is, however, adequately brought out. It is stated in the preface that the book is intended to be both elementary and mathematically sound. The intention is admirably carried out. The author has performed a most valuable service with this well-produced book in making accessible to a wider circle a field of mathematical investigation which is of no less interest to the physicist and engineer than to the mathematician. E. C. S.

Acid-Base Catalysis. By R. P. Bell. [Pp. viii+211.] (The Clarendon Press, Oxford, 1941. Price 12*s.* 6*d.*)

Although catalysis by acids and bases has been studied extensively in this country, as elsewhere, during the past fifty years, this is the first book on the subject by an Englishman. This does not necessarily imply diffidence on the part of potential authors; it has doubtless been conditioned simply by the fact that only in very recent years has it become possible to present anything like a coherent picture of the subject at all. The appearance of the book is therefore well timed, and it will certainly be welcomed by everyone interested in this branch of science.

The balance is fairly held between experiment and theory. The first half of the book deals mainly with experimental material, as may be judged from the fact that twenty tables of experimental figures are included. Almost at the beginning the author introduces the two fundamental topics of salt effects and the modern theory of acids and bases, and in the light of these ideas he then goes on to review and correlate the published data both for aqueous and non-aqueous solutions. The literature is extensive, and in the hands of a less skilful writer this section might easily have degenerated into a mere catalogue of reactions. The author has, however, sifted his material carefully. Much of the older data is necessarily unsuitable for inclusion, but in such cases as seemed to warrant it he has applied to the original figures the numerical corrections required by modern knowledge. The result is a most valuable summary of the best experimental work carried out in this field during the present century.

On the sound basis of established fact thus provided, the author then proceeds to the more theoretical aspects of the subject. The chapters on the Brönsted relation and the mechanism of acid-base catalysis are in every way excellent, and his discussion of the transition state theory of reaction rate is a model of lucid exposition.

It is difficult to commend this book too highly. It will be invaluable to a wide range of readers. G. F. S.

ERRATA.

Ser. 7, vol. xx. September 1935.

P. 427, after line 5 from bottom. Add new paragraph :—

To consider this case, we shall substitute, at $x<0$, $-i$ for i in formula (2.1), leaving (3.1) unaltered, and correspondingly substitute $-i\phi$ for $i\phi$ in (3.4), and x for $|x|$ in the first formula (3.12).

P. 428, in the first formulæ of (4.2) and (4.5). For $|x|$ read x in each case.

P. 430, in the second formula of (4.11). *For* $|\xi|$ *read* ξ.

P. 430, in the second formula of (4.12). For $|x|$ read x.

P. 433, after formula (5.9). Add new paragraph :—

Formulæ (3.4), (4.5) or alternatively (4.11), and (5.5)–(5.9) form in conjunction a set of expressions for the e.m. Whirl, which is valid for $x<0$, as well as for $x>0$, since substituting $-i$ for i in (3.4), and correspondingly in (5.5), would not affect the result. In the case of a stationary wave train changing with time, (4.5) or (4.11) should be replaced by (3.12).

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

XXVI. *The Stresses in an Incomplete Ring subject to Uniform Shear and Compression on its Inner Surface.*

By S. T. Newing, B.Sc., A.R.C.S., D.I.C.*

[Received January 2, 1942.]

I. *Introduction.*

In the evaluation of the stresses in a certain maneton joint, some estimate is required of those arising from normal and tangential forces applied to its inner boundary by the crank pin. Of particular interest is the value of these stresses when the normal and tangential forces are related by the condition for limiting friction at the inner boundary.

The size and shape of the maneton are such that it may be regarded as composed of two parts. One of these is sufficiently massive to be regarded as a rigid base to which the other is attached. For the purpose of analysis this latter part is replaced by an incomplete ring, bounded by circles of radii "a" and "b," which subtend an angle $3\pi/2$ at their common centre. It is the stresses in the incomplete ring with which this investigation is concerned.

The ring is illustrated in fig. 1. The ring $\left(\frac{\pi}{4} \leqslant \theta \leqslant \frac{7\pi}{4}\right)$ is shown completed by the addition of the quadrantal part ABCD. This is introduced so that to it, a system of forces may be applied to equilibrate the forces applied in the other three quadrants. These applied forces have been taken as the uniform radial and tangential forces, "p" and "T" respectively, acting on the boundary $r=b$ as shown in fig. 1; and they represent the forces which the crank pin exerts on the elastic part of the maneton. They are equilibrated over the quadrant AD by a uniform radial (compressive) force "p," and a certain arrangement of tangential tractions over the inner boundary of the quadrant.

II. *Theory and Formulæ.*

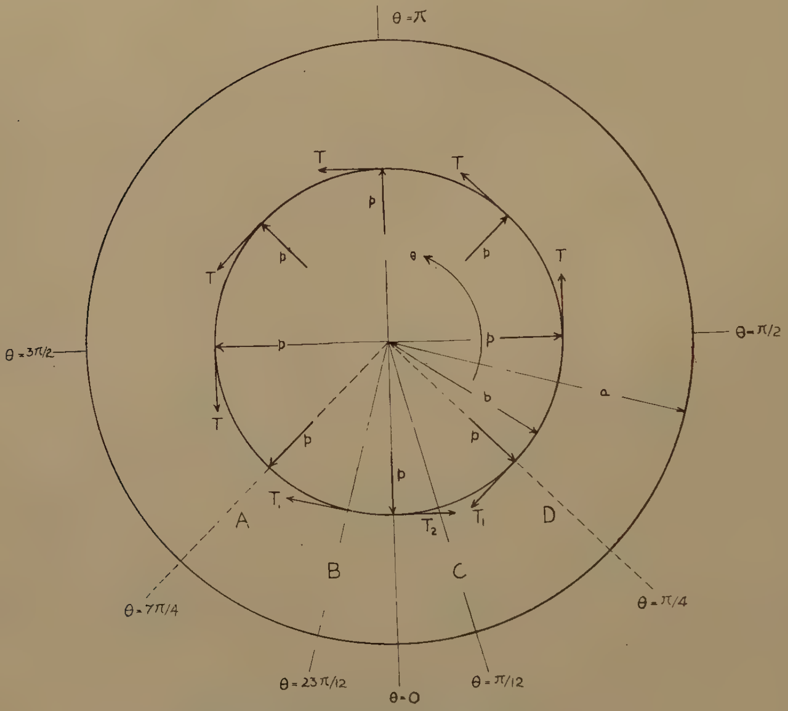
On the usual assumptions of generalized plane stress a solution of the equation $\nabla^4\chi=0$ in polar co-ordinates may be employed to derive expressions for the stresses $\breve{rr}$, $\breve{\theta\theta}$, $\breve{r\theta}$. These expressions contain some terms independent of θ, and others which involve $\sin n\theta$ and $\cos n\theta$, and are as follows :

---

* Communicated by the Author.

$$\begin{aligned}\breve{rr}=2\mathrm{A}_0+\mathrm{B}_0r^{-2}+&\cos\theta\,[2\mathrm{A}_1r-2\mathrm{B}_1r^{-3}-(3+\eta)\mathrm{C}_1r^{-1}/2]\\&+\sin\theta[2\mathrm{A}_1'r-2\mathrm{B}_1'r^{-3}-(3+\eta)\mathrm{C}_1'r^{-1}/2]\\&-\sum_{n=2}^{\infty}\cos n\theta[(n+1)(n+2)\mathrm{A}_nr^n+n(n+1)\mathrm{B}_nr^{-n-2}+n(n-1)\mathrm{C}_nr^{n-2}\\&\qquad+(n-1)(n+2)\mathrm{D}_nr^{-n}]\\&-\sum_{n=2}^{\infty}\sin n\theta[(n+1)(n-2)\mathrm{A}_n'r^n+n(n+1)\mathrm{B}_n'r^{-n-2}+n(n-1)\mathrm{C}_n'r^{n-2}\\&\qquad+(n-1)(n+2)\mathrm{D}_n'r^{-n},\end{aligned}$$

Fig. 1.



Distribution of stresses applied to ring.

$$\begin{aligned}\breve{\theta\theta}=2\mathrm{A}_0-\mathrm{B}_0r^{-2}+&\cos\theta[6\mathrm{A}_1r-2\mathrm{B}_1r^{-3}+(1-\eta)\mathrm{C}_1r^{-1}/2]\\&+\sin\theta[6\mathrm{A}_1'r+2\mathrm{B}_1'r^{-3}+(1-\eta)\mathrm{C}_1'r/2]\\&+\sum_{n=2}^{\infty}\cos n\theta[(n+1)(n+2)\mathrm{A}_nr^n+n(n+1)\mathrm{B}_nr^{-n-2}+n(n-1)\mathrm{C}_nr^{n-2}\\&\qquad+(n-1)(n-2)\mathrm{D}_nr^{-n}]\end{aligned}$$

$$+\sum_{n=2}^{\infty}\sin n\theta[(n+1)(n+2)A_n'r^{n}+n(n+1)B_n'r^{-n-2}+n(n-1)C_n'r^{n-2}$$
$$+(n-1)(n-2)D_n'r^{-n}],$$

$$\overset{\smile}{r\theta}=C_0r^{-2}+\sin\theta[2A_1r-2B_1r^{-3}+(1-\eta)C_1r^{-1}/2]-\cos\theta[2A_1'r-2B_1'r^{-3}$$
$$+(1-\eta)C_1'r^{-1}/2]$$
$$-\sum_{n=2}^{\infty}\cos n\theta[n(n+1)A_n'r^{n}-n(n+1)B_n'r^{-n-2}+n(n-1)C_n'r^{n-2}$$
$$-n(n-1)D_n'r^{-n}]$$
$$+\sum_{n=2}^{\infty}\sin n\theta[n(n+1)A_nr^{n}-n(n+1)B_nr^{-n-2}+n(n-1)C_nr^{n-2}$$
$$-(n-1)D_nr^{-n}].$$

Provided that the given applied stresses $\overset{\smile}{rr}$ and $\overset{\smile}{r\theta}$ on the boundaries of a ring are such that they may be expanded as a Fourier series, then it is possible to obtain the coefficients in the above formulæ for the stresses.

Thus for $\overset{\smile}{rr}$ at $r=b$, there will be a Fourier series which contains a term independent of θ, a series of terms involving $\sin n\theta$, and another set involving $\cos n\theta$. By comparison with the above form, a relation is obtained between A_0 and B_0, together with other relations to be satisfied by the two sets of coefficients A_n, B_n, C_n, D_n and A_n', B_n', C_n', D_n'. Three similar sets of relations are obtained by using the Fourier expansions of $\overset{\smile}{rr}$ at $r=a$, and those for $\overset{\smile}{r\theta}$ at $r=a$ and $r=b$. Thus for each set of four coefficients corresponding to a given value of n, there are four independent equations to determine their value. The coefficients associated with the terms independent of θ are likewise determined.

Commencing with these latter terms, the pair A_0 and B_0, which are associated with $\overset{\smile}{rr}$, are found from the boundary values of that stress as follows:

$$\text{at}\quad r=a,\quad \overset{\smile}{rr}=0,\qquad B_0a^{-2}+2A_0=0,$$
$$\text{at}\quad r=b,\quad \overset{\smile}{rr}=-p,\quad B_0b^{-2}+2A_0=-p,$$

whence $$A_0=pb^2/2(a^2-b^2)$$

and $$B_0=-pa^2b^2/(a^2-b^2).$$

The coefficient C_0 is clearly zero, since the mean value of $\overset{\smile}{r\theta}$ over either boundary is zero.

Passing then to the terms of the first order, it may be noted that, since the total force on each boundary vanishes, the solution is independent of the elastic constant η *. Therefore, as a consequence of the statical equilibrium, the constants C_1 and C_1' are both zero.

---

* Coker and Filon, 'Photo-elasticity,' p. 379.

For the other coefficients, the expressions either for $\overset{\smile}{rr}$ or $\overset{\smile}{r\theta}$ give,

$$\text{at} \quad r=a, \qquad 2A_1a-2B_1a^{-3}=0, \qquad 2A_1'a-2B_1'a^{-3}=0,$$

$$\text{and at} \quad r=b, \qquad 2A_1b-2B_1b^{-3}=0, \qquad 2A_1'b-2B_1'b^{-3}=0;$$

whence

$$A_1=A_1'=B_1=B_1'=0.$$

For the coefficients associated with the terms of order n, there are two sets of four equations. If the Fourier expansion for $\overset{\smile}{r\theta}$ at $r=b$ may be written as $\sum\limits_{n=1}^{\infty} K_n \cos n\theta$, and it will appear later that this is so, then we have the following sets of equations :

$$n(n+1)A_n'a^n-n(n+1)B_n'a^{-n-2}+n(n-1)C_n'a^{n-2} -n(n-1)D_n'a^{-n}=0, \quad . \quad . \quad (2.1)$$

$$n(n+1)A_n'b^n-n(n+1)B_n'b^{-n-2}+n(n-1)C_n'b^{n-2} -n(n-1)D_n'b^{-n}=K_n, \quad . \quad (2.2)$$

$$(n+1)(n-2)A_n'a^n+n(n+1)B_n'a^{-n-2}+n(n-1)C_n'a^{n-2} +(n-1)(n+2)D_n'a^{-n}=0, \quad (2.3)$$

$$(n+1)(n-2)A_n'b^n+n(n+1)B_n'b^{-n-2}+n(n-1)C_n'b^{n-2} +(n-1)(n+2)D_n'b^{-n}=0, \quad (2.4)$$

together with

$$n(n+1)A_na^n-n(n+1)B_na^{-n-2}+n(n-1)C_na^{n-2} -n(n-1)D_na^{-n}=0,$$

$$n(n+1)A_nb^n-n(n+1)B_nb^{-n-2}+n(n-1)C_nb^{n-2} -n(n-1)D_nb^{-n}=0,$$

$$(n+1)(n-2)A_na^n+n(n+1)B_na^{-n-2}+n(n-1)C_na^{n-2} +(n-1)(n+2)D_na^{-n}=0,$$

$$(n+1)(n-2)A_nb^n+n(n+1)B_nb^{-n-2}+n(n-1)C_nb^{n-2} +(n-1)(n+2)D_nb^{-n}=0,$$

for all values of n from two upwards.

It is clear from the form of the second set of equations that $A_n=B_n=C_n=D_n=0$. Moreover, when K_n is known equations (2.1) to (2.4) will determine the values of A_n', B_n', C_n', D_n'.

III. *Determination of the Coefficients* K_n.

The given uniform tangential forces over the surface $r=b$, between $\theta=\pi/4$ and $\theta=7\pi/4$, may be equilibrated by various distributions of force over the part of that surface lying within the quadrant AD. Although there is no unique manner in which equilibrium may be obtained, the stresses in the region $\pi/2<\theta<3\pi/2$, which is of greatest interest, will not depend upon the particular equilibrating system chosen. For, by

St. Venant's principle, any two equipollent loadings applied to the quadrant AD, will produce identical stresses in a region well removed from that quadrant. The manner in which it was decided to equilibrate the applied forces over the incomplete ring, is indicated in fig. 1.

The radial forces over the incomplete ring are equilibrated by assuming the same uniform radial stress, $-p$, to act over the surface $r=b$ of the quadrant AD. This is not the only possible manner in which equilibrium may be realized, but it is the most convenient. The tangential forces are equilibrated by a uniform tangential force T, acting in the clockwise sense (θ decreasing) over those parts of the surface $r=b$ between $\theta=\pi/12$ and $\theta=\pi/4$, and also between $\theta=7\pi/4$ and $\theta=23\pi/12$.

In addition, there is a uniform tangential force T_2, acting in the counter-clockwise sense (θ increasing) over that part of $r=b$ between $\theta=-\pi/12$ and $\theta=\pi/12$.

The values of T_1 and T_2, in terms of T, are obtained from the usual equations of statical equilibrium. By the symmetry of the force system, the total traction over $r=b$ in the direction $\theta=0$ is zero. In the direction perpendicular to this, the condition for zero total force gives

$$2\int_{\pi/4}^{\pi} \mathrm{T}\cos\theta\,.\,b\,d\theta-2\int_{\pi/12}^{\pi/4} \mathrm{T}_1\cos\theta\,.\,b\,d\theta+2\int_{0}^{\pi/12} \mathrm{T}_2\cos\theta\,.\,b\,d\theta=0.$$

In order that the moment of the system about the centre of the ring shall be zero, there is the further condition that

$$2\mathrm{T}\,.\,b\,.\,3\pi/4-2\mathrm{T}_1 b(\pi/4-\pi/12)+\mathrm{T}_2\,.\,b\,.\,\pi/6=0.$$

The two equations which determine T_1 and T_2 are therefore

$$\mathrm{T}/\sqrt{2}+\mathrm{T}_1(1/\sqrt{2}-\sin\pi/12)-\mathrm{T}_2\sin\pi/12=0,$$

and $$9\mathrm{T}-2\mathrm{T}_1+\mathrm{T}_2=0,$$

whence $$\mathrm{T}_1=43{\cdot}7846080\,\mathrm{T},$$

and $$\mathrm{T}_2=78{\cdot}5692160\,\mathrm{T}.$$

The function $f(\theta)=\breve{r\theta}$ at $r=b$, which is to be expanded as a Fourier Series, is therefore defined as follows:

$$\begin{aligned} f(\theta)&=78{\cdot}569\,2160\,\mathrm{T}, && -\pi/12<\theta<\pi/12,\\ f(\theta)&=-43{\cdot}784\,6080\,\mathrm{T}, && \pi/12<\theta<\pi/4 \text{ and } 7\pi/4<\theta<23\pi/12,\\ f(\theta)&=\mathrm{T}, && \pi/4<\theta<7\pi/4. \end{aligned}$$

Owing to its symmetry about $\theta=0$, $f(\theta)$ will have cosine terms only in its expansion, and the coefficient K_n is determined from the formula

$$\pi\mathrm{K}_n=\int_0^{2\pi} f(\theta)\cos n\theta\,d\theta,$$

with $n=2, 3, 4, \ldots$ ad inf. K_0 and K_1 are zero, since the associated integrals vanish in view of the fact that there is statical equilibrium.

TABLE I.

n.	K_n/T.	n.	K_n/T.	n.	K_n/T.
1..	0·00000	25..	0·000000	49..	0·000000
2..	5·21783	26..	0·401372	50..	0·208713
3..	11·63949	27..	1·29237	51..	0·684675
4..	16·86430	28..	2·40919	52..	1·29725
5..	19·07973	29..	3·28962	53..	1·79998
6..	17·73394	30..	3·54679	54..	1·97044
7..	13·62841	31..	3·07738	55..	1·73453
8..	8·43215	32..	2·10804	56..	1·20459
9..	3·87984	33..	1·05813	57..	0·612604
10..	1·04357	34..	0·306931	58..	0·179925
11..	0·00000	35..	0·00000	59..	0·00000
12..	0·00000	36..	0·00000	60..	0·00000
13..	0·00000	37..	0·00000	61..	0·00000
14..	—0·745405	38..	—0·274623	62..	—0·168317
15..	—2·32789	39..	—0·895344	63..	—0·554261
16..	—4·21607	40..	—1·68643	64..	—1·05402
17..	—5·61170	41..	—2·32680	65..	—1·46768
18..	—5·91131	42..	—2·53342	66..	—1·61218
19..	—5·02099	43..	—2·21858	67..	—1·42386
20..	—3·37286	44..	—1·53312	68..	—0·992018
21..	—1·66275	45..	—0·775965	69..	—0·506064
22..	—0·474348	46..	—0·226862	70..	—0·149081
23..	0·000000	47..	0·000000	71..	0·000000
24..	0·000000	48..	0·000000	72..	0·000000

TABLE II.

θ.	$r=12$.	$r=24$.	$r=36$.	$r=48$.	$r=60$.	$r=72$.
90° ..	1·30111	0·84324	1·10538	0·92073	1·06349	0·94703
100° ..	0·92460	1·06120	1·07280	1·02601	0·97907	0·96586
110° ..	0·90581	1·03493	1·05280	1·02108	0·98507	0·97376
120° ..	1·12026	0·93899	1·04003	0·95851	1·02375	0·97874
130° ..	0·95804	1·02678	1·03329	1·01202	0·98971	0·98336
140° ..	0·95308	1·01989	1·02914	1·01175	0·99249	0·98662
150° ..	1·07481	0·96247	1·02522	0·98123	1·01439	0·98673
160° ..	0·96756	1·01777	1·02219	1·00766	0·99212	0·98759
170° ..	0·96623	1·01595	1·02166	1·00801	0·99320	0·98877
180° ..	1·06426	0·96761	1·02243	0·98466	1·01380	0·99004

K_n was evaluated from $n=2$ to 72 inclusive.

In Table I. are recorded the values of K_n/T for the n values 1 to 72.

In Table II. are shown the values of $\sum\limits_{n=1}^{n=r}(K_n \cos n\theta)/T$ for various values of r, corresponding with values of θ chosen at 10° intervals between $\theta=90°$ and $\theta=180°$. The values between $\theta=180°$ and $\theta=270°$ are readily obtained from the fact that $\overset{\smile}{r\theta}$ is an even function of θ.

It will be seen from the entries in Table II., that the series for $\overset{\smile}{r\theta}$ at $r=b$ does not converge very rapidly, and that even when seventy-two terms are summed, there is a mean error of about 2 per cent. in the values in the last column. It was thought that the series to be obtained for the other stresses would converge as slowly as that for $\overset{\smile}{r\theta}$ at the inner boundary. Fortunately this did not prove to be so, and, in fact, the numerical values obtained show that, for the most part, the first twenty-four harmonic terms determined the stresses correct to 1 per cent.

IV. *Evaluation of the Coefficients* A_n', B_n', C_n', D_n'.

Having determined the values of K_n/T, it is possible to evaluate the coefficients for the harmonic term of order n.

Before solving equations (2.1) to (2.4), they are divided, in order, by a^{n-2}, b^{n-2}, a^{n-2}, b^{n-2}, giving

$$n(n+1)A_n'a^2-n(n+1)B_n'a^{-2n}+n(n-1)C_n'-n(n-1)D_n'a^{-2n+2}=0,$$

$$n(n+1)A_n'b^2-n(n+1)B_n'b^{-2n}+n(n-1)C_n'-n(n-1)D_n'b^{-2n+2}=K_nb^{2-n},$$

$$(1-2/n).n(n+1)A_n'a^2+n(n+1)B_n'a^{-2n}+n(n-1)C_n' +(1+2/n).n(n-1)D_n'a^{-2n+2}=0,$$

$$(1-2/n).n(n+1)A_n'a^2+n(n+1)B_n'b^{-2n}+n(n-1)C_n' +(1+2/n).n(n-1)D_n'b^{-2+2}=0.$$

Write

$$n(n+1)A_n'=s_n,$$
$$n(n+1)B_n'=t_n,$$
$$n(n-1)C_n'=u_n,$$
$$n(n-1)D_n'=v_n,$$
$$K_nb^{2-n}=w_n,$$

and these equations become

$$s_na^2-t_na^{-2n}+u_n-v_na^{-2n+2}=0, \quad . \quad . \quad . \quad . \quad . \quad (4.1)$$

$$s_nb^2-t_nb^{-2n}+u_n-v_nb^{-2n+2}=w_n. \quad . \quad . \quad . \quad . \quad . \quad (4.2)$$

$$(1-2/n)s_na^2+t_na^{-2n}+u_n+(1+2/n)v_na^{-2n+2}=0. \quad . \quad . \quad . \quad (4.3)$$

$$(1-2/n)s_nb^2+t_nb^{-2n}+u_n+(1+2/n)v_nb^{-2n+2}=0. \quad . \quad . \quad . \quad (4.4)$$

The value of a was taken as 2·38 units and b as 1·365 units. The coefficients s_n, t_n, u_n, v_n, w_n were evaluated for each n value, and the above set of equations was solved. The usual method was employed for their solution, the number of unknowns being eliminated one at a time until a simple equation in one of them remained. The other three were then determined by substitution, and all four values checked by inserting them on the left-hand sides of equations (4.1) to (4.4). These equations were solved for values of n from 1 to 24, and the values of the four coefficients are shown in Table III.

The formulæ for $\overset{\smile}{rr}$ and $\overset{\smile}{\theta\theta}$ given in Section II. indicate that, in addition to the harmonic terms, there are included others, which are independent of θ.

TABLE III.

n.	s_n/T.	t_n/T.	u_n/T.	v_n/T.
1..	$0{\cdot}00000$	$0{\cdot}00000$	$0{\cdot}00000$	$0{\cdot}00000$
2..	$2{\cdot}67818$	$-1{\cdot}19883\times10^{2}$	$-1{\cdot}13590\times10^{1}$	$4{\cdot}27530\times10^{1}$
3..	$1{\cdot}18832$	$-2{\cdot}33865\times10^{2}$	$-5{\cdot}37000$	$8{\cdot}49574\times10^{1}$
4..	$4{\cdot}79746\times10^{-1}$	$-4{\cdot}20085\times10^{2}$	$-2{\cdot}25558$	$1{\cdot}58107\times10^{2}$
5..	$1{\cdot}66727\times10^{-1}$	$-6{\cdot}63329\times10^{2}$	$-8{\cdot}05971\times10^{-1}$	$2{\cdot}59627\times10^{2}$
6..	$4{\cdot}86897\times10^{-2}$	$-9{\cdot}02527\times10^{2}$	$-2{\cdot}40299\times10^{-1}$	$3{\cdot}66321\times10^{2}$
7..	$1{\cdot}16803\times10^{-2}$	$-1{\cdot}03357\times10^{3}$	$-7{\cdot}54540\times10^{-2}$	$4{\cdot}32850\times10^{2}$
8..	$2{\cdot}21415\times10^{-3}$	$-9{\cdot}56678\times10^{2}$	$-1{\cdot}12500\times10^{-2}$	$4{\cdot}11271\times10^{2}$
9..	$3{\cdot}06871\times10^{-4}$	$-6{\cdot}56958\times10^{2}$	$-1{\cdot}57531\times10^{-3}$	$2{\cdot}88619\times10^{2}$
10..	$2{\cdot}43415\times10^{-5}$	$-2{\cdot}62447\times10^{2}$	$-1{\cdot}26960\times10^{-4}$	$1{\cdot}17401\times10^{2}$
11..	$0{\cdot}00000$	$0{\cdot}00000$	$0{\cdot}00000$	$0{\cdot}00000$
12..	$0{\cdot}00000$	$0{\cdot}00000$	$0{\cdot}00000$	$0{\cdot}00000$
13..	$0{\cdot}00000$	$0{\cdot}00000$	$0{\cdot}00000$	$0{\cdot}00000$
14..	$-1{\cdot}13142\times10^{-7}$	$8{\cdot}66209\times10^{2}$	$5{\cdot}99806\times10^{-7}$	$-4{\cdot}06787\times10^{2}$
15..	$-9{\cdot}75126\times10^{-8}$	$3{\cdot}92323\times10^{3}$	$5{\cdot}19065\times10^{-7}$	$-1{\cdot}85790\times10^{3}$
16..	$-4{\cdot}83245\times10^{-8}$	$1{\cdot}02693\times10^{4}$	$2{\cdot}58165\times10^{-7}$	$-4{\cdot}89918\times10^{3}$
17..	$-1{\cdot}74462\times10^{-8}$	$1{\cdot}96942\times10^{4}$	$9{\cdot}36110\times10^{-8}$	$-9{\cdot}45733\times10^{3}$
18..	$-4{\cdot}96261\times10^{-9}$	$2{\cdot}98083\times10^{4}$	$2{\cdot}66743\times10^{-8}$	$-1{\cdot}43984\times10^{4}$
19..	$-1{\cdot}13013\times10^{-9}$	$3{\cdot}62881\times10^{4}$	$6{\cdot}09323\times10^{-9}$	$-1{\cdot}76211\times10^{4}$
20..	$-2{\cdot}02711\times10^{-10}$	$3{\cdot}48586\times10^{4}$	$1{\cdot}09354\times10^{-9}$	$-1{\cdot}53071\times10^{4}$
21..	$-2{\cdot}64832\times10^{-11}$	$2{\cdot}45231\times10^{4}$	$1{\cdot}43363\times10^{-10}$	$-1{\cdot}20172\times10^{4}$
22..	$-1{\cdot}99599\times10^{-12}$	$9{\cdot}96466\times10^{3}$	$1{\cdot}08262\times10^{-11}$	$-4{\cdot}90240\times10^{3}$
23..	$0{\cdot}00000$	$0{\cdot}00000$	$0{\cdot}00000$	$0{\cdot}00000$
24..	$0{\cdot}00000$	$0{\cdot}00000$	$0{\cdot}00000$	$0{\cdot}00000$

They are $2\mathrm{A}_0+\mathrm{B}_0r^{-2}$ and $2\mathrm{A}_0-\mathrm{B}_0r^{-2}$ respectively. Upon inserting the values already found for A_0 and B_0, these terms are given by

$$2\mathrm{A}_0+\mathrm{B}_0r^{-2}=pb^2\{1-a^2/r^2\}/(a^2-b^2),$$

and

$$2\mathrm{A}_0-\mathrm{B}_0r^{-2}=pb^2\{1+a^2/r^2\}/(a^2-b^2).$$

On substituting the values for a and b, and supposing that the coefficient of limiting friction is 1/5, these terms take the values

$$2\mathrm{A}_0+B_0r^{-2}=2{\cdot}450854\,\mathrm{T}\{1-a^2/r^2\},$$

and

$$2\mathrm{A}_0-\mathrm{B}_0r^{-2}=2{\cdot}450854\,\mathrm{T}\{1+a^2/r^2\},$$

when the equilibrium is limiting. The values of the latter quantity, which appears in the expression for $\overset{\smile}{\theta\theta}$, for various values of a/r between 1 and a/b, are shown in Table IV.

V. *Calculation of* $\overset{\smile}{\theta\theta}$.

The formula for $\overset{\smile}{\theta\theta}$ is

$$\overset{\smile}{\theta\theta}=2\mathrm{A}_0-\mathrm{B}_0r^{-2}+\sum_{n=2}^{\infty}\sin n\theta[(n+1)(n+2)\mathrm{A}_n{}'r^n+n(n+1)\mathrm{B}_n{}'r^{-n-2}$$
$$+n(n-1)\mathrm{C}_n{}'r^{n-2}+(n-1)(n-2)\mathrm{D}_n{}'r^{-n}].$$

TABLE IV.

r.	$(2\mathrm{A}_0-\mathrm{B}_0r^{-2})/\mathrm{T}$.
1·3650	9·90171
1·5000	8·62090
1·7000	7·25453
1·8725	6·40125
2·0000	5·92151
2·2000	5·31917
2·3800	4·90171

TABLE V.

θ.	$m=12$.	$m=24$.	$\overset{\smile}{\theta\theta}/\mathrm{T}$.
90° ..	−7·66760	−7·66386	−2·76215
100° ..	−8·76892	−8·77147	−3·86976
110° ..	−9·24226	−9·24180	−4·34009
120° ..	−9·37166	−9·37050	−4·46879
130° ..	−8·92180	−8·92334	−4·02163
140° ..	−7·68693	−7·68618	−2·78447
150° ..	−6·01926	−6·01890	−1·11719
160° ..	−4·29363	−4·29569	0·60602
170° ..	−2·34418	−2·34212	2·55959
180° ..	0·00000	0·00000	4·90171

Now

$$(n+1)(n+2)\mathrm{A}_n{}'=(1+2/n)s_n,$$
$$n(n+1)\mathrm{B}_n{}'=t_n,$$
$$n(n-1)\mathrm{C}_n{}'=u_n,$$
$$(n-1)(n-2)\mathrm{D}_n{}'=(1-2/n)v_n;$$

so that

$$\overset{\smile}{\theta\theta}=2\mathrm{A}_0-\mathrm{B}_0r^{-2}+\sum_{n=2}^{\infty}\sin n\theta[(1+2/n)s_nr^n+t_nr^{-n-2}+u_nr^{n-2}$$
$$+(1-2/n)v_nr^{-n}].$$

With the values of s_n, t_n, u_n, v_n already determined in Section IV. the value of $\overset{\smile}{\theta\theta}$ can be determined up to and including the twenty-fourth harmonic term.

The value of $\overset{\smile}{\theta\theta}$ at the boundary $r=a$ was first calculated, and it was found that the sum of the first twenty-four harmonic terms differed by less than 1 part in 2000 from the sum of the first twelve harmonic terms. Values of $\overset{\smile}{\theta\theta}$ were calculated at intervals of 10° from 90° to 180°. To obtain the values corresponding with θ values between 180° and 270°, it is necessary only to reverse the sign of the harmonic terms, leaving $2A_0-B_0a^{-2}$ unchanged.

If, at $r=a$,

$$\overset{\smile}{\theta\theta}=2A_0-B_0a^{-2}+\sum_{n=1}^{m}\sin n\theta[(1+2/n)s_na^n+t_na^{-n-2}+\ldots, \text{etc.}],$$

Table V. gives the values of $\overset{\smile}{\theta\theta}/T$ and the values of the sum in the above expression for $m=12$ and $m=24$, when divided by T.

These values are illustrated graphically in fig. 2.

At the inner boundary, it was found that more than twenty-four terms were required to give a sum with a comparable degree of convergence. For large values of n, it was found possible to calculate the coefficient of $\sin n\theta$ in the above formula, without the considerable labour involved in the solution of equations (4.1) to (4.4). From Table III. it is clear that, for values of n greater than 24, the coefficients s_n and u_n are likely to be very small. Calculation verifies that the terms in equations (4.2) and (4.4), which involve s_n and u_n, are negligible in comparison with those arising from t_n and v_n. Accordingly t_n and v_n may be determined by neglecting s_n and u_n in equations (4.2) and (4.4), and then by substitution of these values in (4.1) and (4.3), an estimate of s_n and u_n can be obtained. It can then be verified that the terms neglected in (4.2) and (4.4) are, in fact, small enough to justify this procedure. It transpires that this approximate procedure gives values for the coefficients of $\sin n\theta$ appearing in the formula for the stress $\overset{\smile}{\theta\theta}$, which for $n>12$, agree with the exact values to four significant figures and for $n>17$ to six significant figures. Therefore, as it was required to include many harmonics above the twenty-fourth in order to find $\overset{\smile}{\theta\theta}$ at $r=b$, this procedure was adopted in finding the contribution made to $\overset{\smile}{\theta\theta}$ by the harmonics from $n=24$ to $n=72$.

In addition to giving very accurate values for the coefficients of $\sin n\theta$ when n exceeds 17, it is possible also to use the approximate form for the coefficient of $\sin n\theta$ to check the values found by the exact procedure in all but the first few cases.

Neglecting, therefore, the terms involving s_n and u_n in (4.2) and (4.4), these equations take the form

$$-t_nb^{-2n}-v_nb^{-2n+2}=w_n,$$

and

$$t_nb^{-2n}+(1+2/n)v_nb^{-2n+2}=0,$$

whence

$$v_n=nb^{2n-2}w_n/2=nb^nK_n/2, \quad \ldots \ldots \quad (5.1)$$

and

$$t_n=-(1+n/2)b^{2n}w_n=-(1+n/2)b^{n+2}K_n. \quad \ldots \quad (5.2)$$

Neglecting the contributions made to the coefficient of $\sin n\theta$ by s_n and u_n, this quantity becomes

$$t_nb^{-n-2}+(1-2/n)v_nb^n=-2K_n, \quad \ldots \ldots \quad (5.3)$$

on substituting for t_n and v_n their values given in (5.1) and (5.2). Thus, after a certain value m of n,

$$\sum_{n=m}^{n=\infty} \sin n\theta[(1+2/n)s_nb^n+t_nb^{-n-2}+u_nb^{n-2}+(1-2/n)v_nb^{-n}]$$
$$=-\sum_{n=m}^{\infty} 2\mathrm{K}_n \sin n\theta,$$

TABLE VI.

n.	x_n (exact).	x_n (approx.).	n.	x_n (exact).	x_n (approx.).
1..	0·00000	0·00000	13..	0·00000	0·00000
2..	−35·9113	−10·4357	14..	1·49087	1·49081
3..	−40·5098	−23·2790	15..	4·65587	4·65580
4..	−43·8773	−33·7286	16..	8·43220	8·43214
5..	−43·1987	−38·1600	17..	11·2234	11·2234
6..	−37·5450	−35·4679	18..	11·8226	11·8226
7..	−28·0016	−27·2568	19..	10·0420	10·0420
8..	−17·0489	−16·8643	20..	6·74572	6·74572
9..	− 7·78628	− 7·75968	21..	3·32550	3·32550
10..	− 2·09103	− 2·08714	22..	0·948697	0·948697
11..	0·00000	0·00000	23..	0·00000	0·00000
12..	0·00000	0·00000	24..	0·00000	0·00000

TABLE VII.

θ.	$m=12$.	$m=24$.	$m=36$.	$m=48$.	$m=60$.	$m=72$.
90° ..	17·5264	17·3774	17·8477	17·8254	17·8386	17·8298
100° ..	18·7550	18·9720	19·1567	19·2446	19·2223	19·1469
110° ..	19·9167	19·8115	19·6748	19·6029	19·6134	19·6679
120° ..	19·2163	19·1842	19·1937	19·1891	19·1918	19·1900
130° ..	17·5615	17·6562	17·7423	17·7843	17·7747	17·7397
140° ..	15·5112	15·4460	15·3724	15·3346	15·3414	15·3708
150° ..	12·2016	12·1926	12·1952	12·1939	12·1947	12·1942
160° ..	8·31550	8·37727	8·43707	8·46667	8·46045	8·43643
170° ..	4·40465	4·34976	4·29270	4·26394	4·26952	4·29236
180° ..	0·00000	0·00000	0·00000	0·00000	0·00000	0·00000

and for the present investigation, m may be taken to be 17. In Table VI. are shown the values of the coefficients of sin $n\theta$ from $n=1$ to $n=24$; the first column shows the values obtained by the exact method, which may be compared directly with those in the second column, determined by the approximate expression (5.3).

The coefficient of sin $n\theta$ is denoted by x_nT.

Table VII shows the values of $\sum_{n=1}^{m} x_n \sin n\theta$ for values of θ at 10° intervals between 90° and 180°, for various m values.

The values of $\overset{\smile}{\theta\theta}/\mathrm{T}$ given by the last column in Table VII. together with the term $2\mathrm{A}_0-\mathrm{B}_0b^{-2}$, are given in Table VIII. and illustrated graphically in fig. 2.

An expression for $\overset{\smile}{\theta\theta}$ at $r=a$ may also be obtained by the approximate method already described. In this instance, s_n and u_n cannot be neglected altogether, but they may be evaluated by substituting in (4.1) and (4.3) the values of t_n and v_n, obtained by neglecting them in (4.2) and (4.4). This procedure gives

$$s_n=nt_na^{-2n-2}+(n+1)v_na^{-2n},$$
$$u_n=(1-n)t_na^{-2n}-nv_na^{-2n+2},$$

TABLE VIII.

θ.	$\overset{\smile}{\theta\theta}/\mathrm{T}$.
90° ..	27·7315
100° ..	29·0486
110° ..	29·5696
120° ..	29·0917
130° ..	27·6414
140° ..	25·2575
150° ..	22·0959
160° ..	18·3381
170° ..	14·1941
180° ..	9·90171

TABLE IX.

$\overset{\smile}{\theta\theta}/\mathrm{T}$.

θ.	$r=1{\cdot}7000$.	$r=1{\cdot}8725$.	$r=2{\cdot}0000$.
90° ..	12·899	7·978	4·816
100° ..	13·108	7·798	4·403
110° ..	13·281	7·691	4·256
120° ..	13·101	7·504	4·140
130° ..	12·450	7·301	4·030
140° ..	11·723	7·075	4·096
150° ..	10·936	6·924	4·474
160° ..	9·750	6·794	5·043
170° ..	8·385	6·633	5·555
180° ..	7·255	6·410	5·922

whence, on substituting for t_n and v_n their values as given by (5.1) and (5.2), the coefficient of $\sin n\theta$ is found to be given by

$$(1+2/n)s_na^n+t_na^{-n-2}+u_na^{n-2}+(1-2/n)v_na^{-n}$$
$$=(b/a)^n\,.\,2\mathrm{K}_n[n-(n+2)(b/a)^2]. \qquad (5.4)$$

Thus, when $n=2$, the value of the coefficient of $\sin n\theta$ as calculated from the R.H.S. of (5.4) is $-6{\cdot}52729\times10^{-5}$.

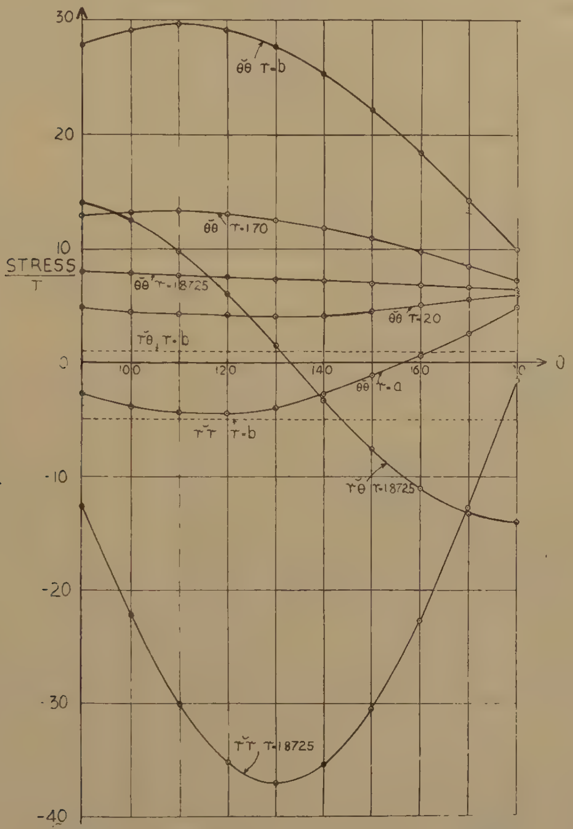
The value obtained by direct solution of the equations (4.1) to (4.4) and substitution on the L.H.S. of (5.4) is $-6{\cdot}52743\times10^{-5}$.

The coefficients of $\sin n\theta$ in $\overset{\smile}{\theta\theta}$ at $r=a$, from $n=24$ to $n=36$, were calculated from equation 5.4, and in no case did the coefficient exceed $1\times10^{-5}\mathrm{T}$ in absolute value. Accordingly, the contribution to $\overset{\smile}{\theta\theta}$ at $r=a$, by harmonics of order higher than the twenty-fourth, were neglected.

since they leave the fifth decimal place unaffected in those values which have appeared already in Table V.

Having investigated the variation of $\breve{\theta\theta}/\mathrm{T}$ with θ for the inner and outer boundaries, it was decided to determine the nature of that variation for other values of r. The values chosen were $r=1{\cdot}7000$, the mean radius

Fig. 2.



$r=1{\cdot}8725$, and $r=2{\cdot}0000$. If $x_n(r)$ denote the coefficient of $\sin n\theta$ in $\theta\theta$, for a certain value of r, then

$$x_n(r)=(1+2/n)s_nr^n+t_nr^{-n-2}+u_nr^{n-2}+(1-2/n)v_nr^{-n},$$

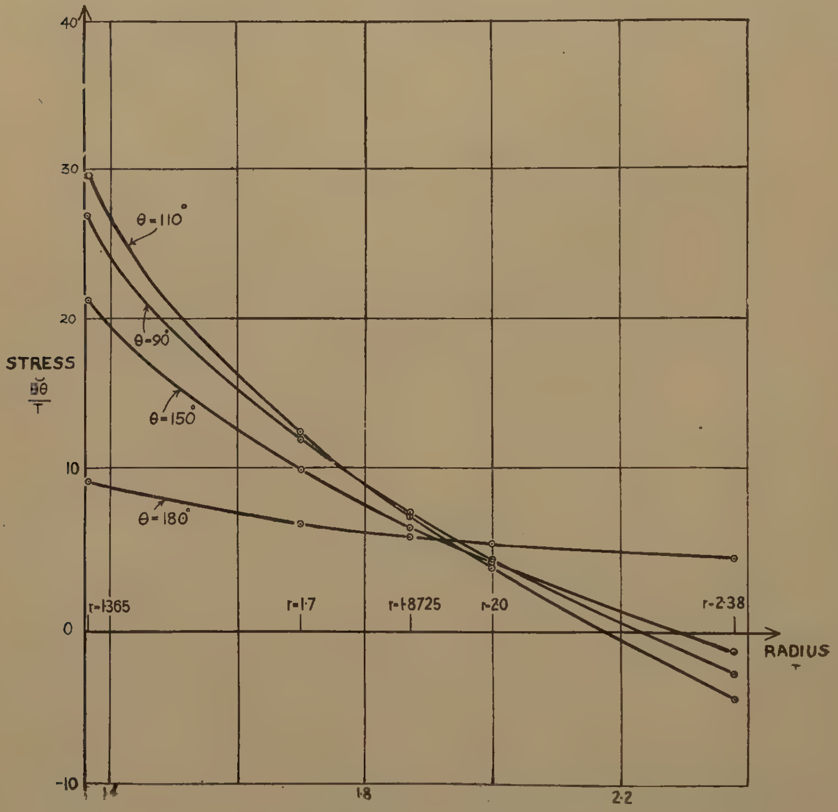
and

$$\breve{\theta\theta}=2\mathrm{A}_0-\mathrm{B}_0r^{-2}+\sum_{n=1}^{\infty}x_n(r)\sin n\theta.$$

For these calculations, only the first twenty-four harmonic terms were included. The extent to which the series had converged at this stage is intermediate between that for the values $r=a$ and b. For most of the calculations, the total contributions of the harmonics between the twelfth and twenty-fourth, affected the third decimal place, and so the values of $\widetilde{\theta\theta}$ shown in Table IX. are given only to the nearest third place of decimals.

These same values are also illustrated in fig. 2. In fig. 3, is shown the variation of $\widetilde{\theta\theta}/T$ with r for certain values of θ, 90°, 110°, 150°, 180°, the values plotted being obtained from Tables V., VIII., and IX.

Fig. 3.



Variation of hoop stress with *r*.

VI. *Evaluation of Radial Stress at Mean Radius.*

Referring to the formula given in Section II. for the radial stress, we have

$$\widetilde{rr}=2A_0+B_0r^{-2}-\sum_{n=2}^{\infty}\sin n\theta[(n+1)(n-2)A_n'r^n+n(n+1)B_n'r^{-n} \quad 2$$
$$+n(n-1)C_n'r^{n-2}+(n-1)(n+2)D_n'r^{-n}],$$

from which it is clear, that $\overset{\smile}{rr}$ possesses the same type of symmetry (odd) about $\theta=0$, as does $\overset{\smile}{\theta\theta}$, in so far as the harmonic terms are concerned.

Now
$$(n+1)(n-2)A_n'=(1-2/n)s_n,$$
$$n(n+1)B_n'=t_n,$$
$$n(n-1)C_n'=u_n,$$
$$(n-1)(n+2)D_n'=(1+2/n)v_n,$$
$$2A_0+B_0r^{-2}=-1{\cdot}50855\,T$$

at the mean radius, when the equilibrium is limiting.

Thus for that value of r (1·8725),

$$\overset{\smile}{rr}=-1{\cdot}50855\,T-\sum_{n=2}^{\infty} y_n \sin n\theta,$$

where $$y_n=(1-2/n)s_nr^n+t_nr^{-n-2}+u_nr^{n-2}+(1+2/n)v_nr^{-n}$$

and $$r=1{\cdot}8725.$$

TABLE X.

θ.	$\sum_{n=2}^{12}(y_n \sin n\theta)/T$.	$\sum_{n=2}^{24}(y_n \sin n\theta)/T$.	$\overset{\smile}{rr}/T$.
90° ..	−11·194	−11·148	−12·657
100° ..	−20·645	−20·690	−22·199
110° ..	−28·642	−28·626	−30·135
120° ..	−33·682	−33·670	−35·179
130° ..	−35·496	−35·519	−37·028
140° ..	−33·917	−33·903	−35·412
150° ..	−29·011	−29·008	−30·517
160° ..	−21·219	−21·233	−22·742
170° ..	−11·250	−11·238	−12·747
180° ..	0·000	0·000	− 1·509

The values of $\overset{\smile}{rr}$ were calculated by summing, first the first twelve harmonics, and secondly the first twenty-four harmonics. The difference between the contributions to $\overset{\smile}{rr}$ of these two sets of harmonics is less than 1 part in 100, and is illustrated in Table X., in which the relevant numbers are given to the nearest third decimal place. The values of $\overset{\smile}{rr}/T$ in the last column are illustrated graphically in fig. 2.

VII. *Evaluation of Shear Stress at Mean Radius.*

The formula for the shear stress within the ring according to Section II. is

$$\overset{\smile}{r\theta}=-\sum_{n=2}^{\infty}\cos n\theta\big[n(n+1)A_n'r^n-n(n+1)B_n'r^{-n-2}$$
$$+n(n-1)C_n'r^{n-2}-n(n-1)D_n'r^{-n}\big].$$

On substituting for the coefficients, this expression becomes

$$\overset{\smile}{r\theta}=-\sum_{n=2}^{\infty} f_n(r)\cos n\theta,$$

where

$$f_n(r)=s_n r^n-t_n r^{-n-2}+u_n r^{n-2}-v_n r^{-n}.$$

This formula, for $\overset{\smile}{r\theta}$, shows that the shear stress possesses even symmetry about $\theta=0$, which enables the shear to be written down at $\theta=-\alpha$ when the value corresponding with $\theta=\alpha$ is known.

The values of $f_n(r)$ for the mean radius $r=1{\cdot}8725$ were calculated from $n=1$ to $n=24$, and the shear stresses were calculated for θ values varying by 10° intervals between $\theta=90°$, and $\theta=180°$. The sum of the first twenty-four harmonic terms differs by less than 1 part in 1000 from the sum of the first twelve harmonic terms. The values of $\overset{\smile}{r\theta}/\mathrm{T}$ to three places of decimals are taken from the last column of Table XI. and are illustrated graphically in fig. 2.

TABLE XI.

θ.	$\sum_{n=2}^{12}(f_n\cos n\theta)/\mathrm{T}$.	$\sum_{n=2}^{24}(f_n\cos n\theta)/\mathrm{T}$.
90°	13·881	13·892
100°	12·450	12·451
110°	9·738	9·731
120°	5·977	5·983
130°	1·509	1·510
140°	− 3·339	− 3·342
150°	− 7·670	− 7·666
160°	−11·175	−11·177
170°	−13·309	−13·311
180°	−14·131	−14·128

VIII. *Discussion of Results.*

(a) *Cross-Radial Stress.*

The cross-radial, or hoop, stress $\overset{\smile}{\theta\theta}$, reaches its maximum value, of about 30T, at the inner edge, for $\theta=110°$, and at the outer edge reaches its minimum value, of about $-4{\cdot}5$T, at $\theta=120°$. The values of $\overset{\smile}{\theta\theta}$ at the inner edge are positive, and imply that, for that direction, the material is in tension in that region; whereas, at the outer edge, up to $\theta=157°$, the values are negative, and imply compression in that region. Beyond $\theta=157°$, the stress becomes positive again and the stress is tensile. At the mean radius there is a tensile stress which varies only slightly with angular distance, and is, in fact, very little different from the value which it would have in the absence of the tangential forces on the inner surface of the ring. This value is, of course, given by the terms independent of θ in the formula for $\overset{\smile}{\theta\theta}$. In the quadrant $180° < \theta < 270°$ the values of $\overset{\smile}{\theta\theta}$ are obtained as already explained, by changing the sign of the harmonic terms. At the inner edge, the cross-radial stress will diminish, becoming negative at 204°, and remaining negative for the rest of the quadrant, attaining a minimum value of about $-10\,T$ at 250°. Thus there will be tensile stress from 180° to 204°, and compressive stress over the rest of the quadrant. At the outer edge, the stress is tensile everywhere in this quadrant, and reaches a maximum value of about $-14{\cdot}5\,T$ at $\theta=240°$.

At $r=1{\cdot}7$ and $2{\cdot}0$, the cross-radial stress is tensile over both quadrants considered, and the range of variation in the actual values is much less marked than for either edge. At $r=1{\cdot}7$ the stress has a maximum value at $\theta=110°$, of about 13 T, and a minimum value, of about T, at $\theta=250°$, whereas at $r=2{\cdot}0$, there is a very broad minimum, of about 4 T, at $\theta=130°$, and a similar maximum, of about 8 T, at $\theta=230°$.

Fig. 3 illustrates the general decrease in cross-radial stress for a given value of θ as r increases, the most marked decrease corresponding with $\theta=110°$. For r values between $1{\cdot}90$ and $1{\cdot}95$ the cross-radial stress (tensile) has nearly the same value, about 6 T, for all angular values.

(b) *Radial Stress.*

At the mean radius, the radial stress is compressive from $\theta=90°$ to $\theta=180°$, and tensile from about $\theta=180°$, to $\theta=270°$, reaching a minimum value, of about $-37\,T$, at 130° and a maximum value, of about 35 T, at 230°. Since $\overset{\smile}{rr}$ is negative at $r=b$, and zero at $r=a$, it is probable that the radial stress is everywhere compressive in the quadrant 90° to 180°, and for all r values too. In the quadrant 180° to 270°, except for a minute region near the former value, the stress is everywhere tensile.

(c) *Shear Stress.*

The shear stress at the mean radius is positive, *i. e.* is directed in the counterclockwise sense from $\theta=90°$, to $\theta=133°$, becoming zero at the latter value. It is negative from $\theta=133°$ to $\theta=227°$, *i. e.* is directed in the clockwise sense, and after passing through zero at the latter value, is positive for the rest of that quadrant. The values of the shear stress are symmetrical about the ordinate at $\theta=180°$.

XXVII. *The Variation of Viscosity of Liquids with Temperature.*

By M. K. SRINIVASAN and B. PRASAD *.

[Received May 27, 1940.]

THE variation of viscosity of liquids with temperature has been the subject of extensive study. A good account of the earlier work, which consisted mainly of empirical equations, is given in Hatschek's 'The Viscosity of Liquids' (1). Recent work has been reviewed by Ferguson (2), Burgers (3) and Jaeger (4). These surveys do not include a comparative study of the various equations from the point of view of their applicability to known data on organic and inorganic liquids. Such a study should, besides ascertaining the relative merits of the various formulæ for practical purposes, serve to verify the validity of the theoretical assumptions made in evolving them. In this paper a comparison of this nature has been made and the assumptions in some equations tested. Also some very recent work has been examined.

In 1912, Porter (5) very nearly suggested a formula : $F\left(\frac{1}{\eta}\right)=A+B/T$ where $F\left(\frac{1}{\eta}\right)$ is a function of fluidity and A and B are constants. But he did not say what function of fluidity $F\left(\frac{1}{\eta}\right)$ was. Guzman (6), Kendall (7), Raman (8), Dunn (9), Andrade (10) and Sheppard (11) independently suggested in different forms an equation of the type $\eta=Ae^{b/T}$, where A and b are constants. In arriving at the equation, Raman suggested that transfer of momentum in liquids did not take place as in gases only but both as in gases and in solids. He further expected the value of A for any liquid to be of the same order as the viscosity of its vapour. The credit for the equation, however, justly goes to Andrade, who has worked out a mechanism of liquid viscosity from which the equation follows as a consequence (12). Andrade suggests that transfer of momentum in liquids (unlike gases) takes place by molecules of adjacent layers coming in close contact for some time. By utilizing the Boltzmann distribution formula for the relation between temperature and the number of molecules possessing a certain potential energy for the transfer of momentum by temporary union, Andrade deduces the equation $\eta=Ae^{b/T}$.

---

* Communicated by the Authors.

By considering the effect of temperature on the distance between the molecules, he modifies the above equation thus : $\eta v^{1/3} = Ae^{c/vT}$. According to him the modified formula should be applicable to simple as well as associated liquids, because " viscosity effect is in essence due to a kind of association governed by the same temperature law as ordinary association."

The mechanism of transfer of momentum by transfer of material particles was rejected by Andrade and others primarily because viscosity of liquids, instead of increasing, decreases with temperature. Curiously enough, gases under high pressure also behave like liquids (13). It is quite likely that when molecules are brought very close to one another, be it due to compression or liquefaction of a gas, the transfer of momentum is not only due to a transfer of material particles but to actual contact also. In case of liquids also, as the temperature rises, the transfer of momentum due to transfer of material particles will become more important. How far this will affect the first equation of Andrade it is difficult to say. In view of this, as well as the fact that this effect will be in the opposite direction to the one due to the increase in volume with rise in temperature, the simple equation does not appear to need modification. Actually the simple equation gives good results (14). In Table I. are given results which show that the simple equation is as good as the modified equation, the deviations being almost the same. The range $\left(\frac{\eta_{\text{maximum}}}{\eta_{\text{minimum}}}\right)$ over which the simple equation has been applied is given in parentheses, and that over which the modified equation has been applied, outside. Purposely those unassociated liquids have been chosen from

TABLE I.

Liquid.	Range.	Max. % difference (simple) equation.		Max. % difference (modified) equation.	
		+	−	+	−
Ethylene bromide	3·9 (4·6)	1·9	2·3	1·4	2·3
Ethylene chloride	2·7 (2·6)	0·8	0·8	0·6	1·0
Carbon tetrachloride	2·5 (2·5)	1·3	1·3	0·6	1·1
Isobutyl bromide	2·6 (2·5)	0·9	1·0	1·0	0·4
Isobutyl iodide	3·2 (4·4)	1·0	1·2	0·9	1·9
Propylene bromide	4·9 (4·9)	1·7	2·5	1·5	2·8
Isobutylene bromide	6·5 (6·4)	4·6	4·4	2·3	4·6
Acetic anhydride	4·3 (4·0)	1·6	1·3	1·0	2·1
Propionic anhydride	6·3 (5·9)	2·7	2·4	1·7	3·1

Andrade's paper where temperature variation of viscosity is not well represented by either of the equations.

Associated liquids may be divided into two classes: those such as acetic acid, whose association does not vary appreciably with temperature in the experimental range, and those such as water, whose association changes a good deal with temperature. So far as the first group is concerned there is no reason why Andrade's simple equation should not hold. But there is no doubt that association cannot be fully explained by merely assuming that statistically a fraction of molecules are held together due to their coming in contact with one another during the process of transfer of momentum. Hence neither of the equations of Andrade should apply to associated liquids if there is a rapid change of association with temperature. In the case of water the maximum percentage difference goes up to $+27{\cdot}1$ and $-14{\cdot}9$, in the case of ethyl alcohol up to $+4{\cdot}5$ and $-4{\cdot}9$, and in the case of propyl alcohol up to $+2{\cdot}2$ and $-2{\cdot}1$, with the modified equation. Even with the simple equation the viscosity of ethyl alcohol over the same range can be represented by $\log \eta = \bar{5}{\cdot}7182 + \frac{685{\cdot}5}{\mathrm{T}}$, with an error of $+4{\cdot}2$ and $-5{\cdot}7$, and that of propyl alcohol by $\log \eta = \bar{5}{\cdot}1970 + \frac{920}{\mathrm{T}}$, with an error of $+4{\cdot}4$ and $-4{\cdot}4$. So far as associated liquids are concerned none of these equations represents the variation of viscosity with temperature satisfactorily. There is ample justification for believing that if a liquid does not follow the simple equation it is associated. Its association can be calculated on certain assumptions (15). Working from the point of view of the rate of a process in which matter rearranges by surmounting a potential barrier, Eyring(16) derives Andrade's and Batschinski's equations. Lewis (17) derives Andrade's equation from entirely different considerations.

Silverman (18) starts with Maxwell's equation $\eta = \mathrm{E}t$, where E is the elasticity modulus and t the relaxation time of elastic force. He assumes some structure in the liquid state that breaks down and builds up again. This structure he attributes to the formation of temporary associated groups. The relaxation time, according to his view, comes out to be the *average time an associated group exists*. From the law of collision and the Boltzmann distribution law he finds that $t = \frac{k'v}{\sqrt{\mathrm{T}}} e^{\mathrm{Q/RT}}$, where k' and Q are constants and v is the specific volume. He shows that $\mathrm{E} = \frac{3}{\beta}$, where β is the compressibility, and further assumes that $\beta = \frac{3}{\mathrm{A}} e^{c\mathrm{T}}$, where A and c are constants, and thus derives the following equations:

$$\eta = \frac{k'v}{\sqrt{\mathrm{T}}} e^{\mathrm{Q'/RT}}, \quad \ldots \ldots \ldots \quad (1)$$

$$\eta = \frac{k''v}{\sqrt{T}} e^{\frac{Q''}{RT} - cT} . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The second equation takes into account the variation of elasticity with temperature. The first does not. Both the equations represent unassociated organic liquids well. The accuracy reached is, however, a little less than in the case of Andrade's simple equation.

	Temp. range ° C.	Max. % difference.		
		Andrade (simple).	Silverman (simple).	Silverman (modified).
Octane	0–120	1·1	1·7	1·85
Benzene	0– 80	2·1	2·1	2·88

When the equations are applied to gallium, potassium and sodium, it is found that c cannot be given a positive value. With a slight negative value for c or with $c=0$, the variation of viscosity is well represented, as with any of Andrade's equations. The results can be seen in Table II. It is impossible for c to have a negative value.

Macleod (19) has built up a theory of viscosity of liquids on the simple assumption that the viscosity of a liquid is an inverse function of its free space per unit volume. His formula is $\eta_t(x_t)^A = k$, where η_t is viscosity at temperature t, x_t is the free space per unit volume at the same temperature and A is the degree of association. x_t is calculated on the assumption that when a simple liquid increases in volume on heating, the increase in volume goes completely towards the increase in free space. The equation represents unassociated organic liquids well, but fails in the case of liquid gallium whose viscosity has been measured over a wide range of temperature. In a later paper (20) he modifies his original equation to $\eta = \frac{KM_0\alpha}{x}$, where K is a constant for all liquids, M_0 is the molecular weight of the substance in the vapour state and α the degree of association. α varies with temperature. In recent years Macleod has tried to derive equations of the exponential type based on his original conceptions of viscosity. In one of his papers (21) he substitutes $v-b$ for x in the equation $\eta = \frac{KM_0\alpha}{x}$, where v is the specific volume and b van der Waals' constant corresponding to one gram of the liquid. He assumes (1) that M, the molecular weight in the liquid state, is an exponential function of temperature and specific volume ($M = M_0\alpha = M_0 e^{c'/vT}$), (2) that in the van der Waals' equation, $(p+\pi)(v-b) = nRT$ (where π is internal pressure),

p is negligible in comparison with π, and (3) that π is an exponential function of T and v ($\pi = \pi_0 e^{c''/v\mathrm{T}}$). He then substitutes $v-b$ for x and $\mathrm{M}_0 e^{c'/v\mathrm{T}}$ for $\mathrm{M}_0\alpha$ in the equation $\eta = \dfrac{\mathrm{KM}_0\alpha}{x}$, thus obtaining

$$\eta = \frac{\mathrm{B}}{\mathrm{T}} e^{\frac{c'+c''}{v\mathrm{T}}} \quad \text{or} \quad \eta = \frac{\mathrm{B}}{\mathrm{T}} e^{c/v\mathrm{T}},$$

where B and c are constants.

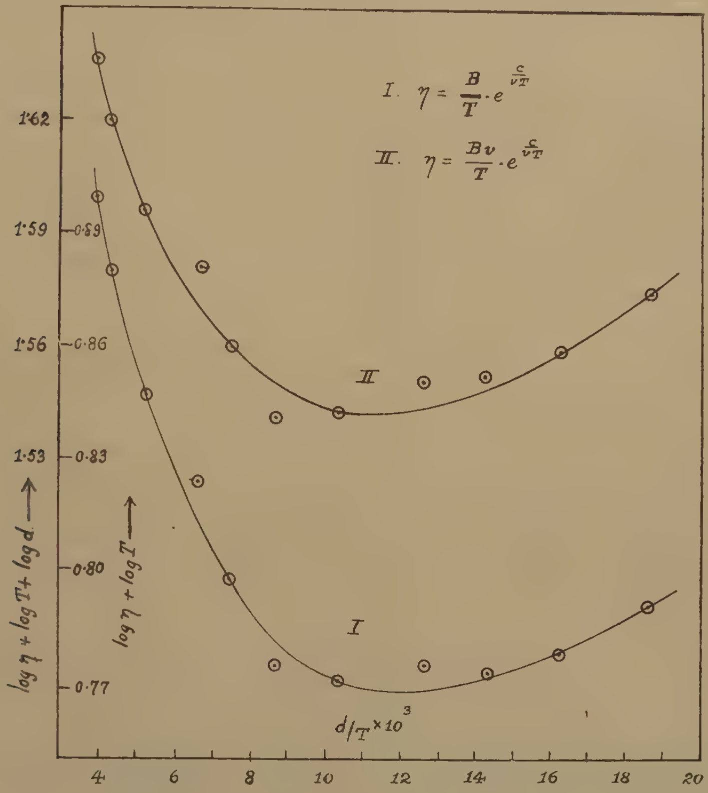
The mistakes that Macleod has made in deriving the equation are: (1) that he puts free space in v c.c. of liquid as equal to $v-b$. If the value of b $\left(i.\,e.\ \dfrac{v_c}{3} \text{ or } \dfrac{v_c}{4}\right)$ calculated from the mass of the liquid from critical constants or from critical density is used, the calculated free space, though of the same order, is far from being the same as the free space per gram ($\overset{s}{f}_t$) calculated from his original idea $\overset{s}{f}_t = \overset{s}{f}_0 + (v_t - v_0)$. (2) Even if the free space per v c.c. is equal to $v-b$, he is not justified in writing $x = v-b$, as x is the free space per unit volume of the liquid and it is really equal to $\dfrac{v-b}{v}$. Putting $\dfrac{v-b}{v}$ for x (instead of $v-b$) and making the same assumptions that Macleod has made, the final equation assumes the form $\eta = \dfrac{\mathrm{B}v}{\mathrm{T}} e^{c/v\mathrm{T}}$ instead of $\eta = \dfrac{\mathrm{B}}{\mathrm{T}} e^{c/v\mathrm{T}}$. So far as octane is concerned the calculated values of viscosity from $\eta = \dfrac{0{\cdot}2932}{\mathrm{T}} e^{563/v\mathrm{T}}$ as well as $\eta = \dfrac{0{\cdot}1246v}{\mathrm{T}} e^{916{\cdot}5/v\mathrm{T}}$ agree with the experimental values within one per cent. But both the equations fail in the case of gallium, as the following figures indicate.

Macleod (22), in another paper, starts with the original equation of Andrade, $\eta = \dfrac{4\nu m}{3\sigma}$ (where ν is the frequency of vibration, m the molecular weight, σ the mean distance between the centres of the molecules), which the latter applied to the viscosity at the melting point only. Macleod applies the equation to all temperatures by taking into account the change in ν and σ with temperature. If a small sphere vibrates in a cylinder of length l having the same cross-section as the sphere, the frequency of vibration will be given by the equation

$$\nu = \frac{\pi r^2 c}{2\pi r^2 h} = \frac{\text{cross-section} \times c}{2 \times \text{free space}},$$

where c is the velocity of the sphere. In liquids, every molecule vibrates within the free space at its disposal. Further, in the case of liquids, he considers the molecules to be spherical and their size independent of temperature. If v_0 is the specific volume when the free space is zero, then the free space per molecule at any temperature can be shown to be $\dfrac{Mv_0}{N} \times \dfrac{v-v_0}{v_0}$, where v is the specific volume at that temperature. The cross-section of the molecules can also be written as $K\left(\dfrac{Mv_0}{N}\right)^{2/3}$.

Fig. 1.



Substituting the free space and the cross-section in the above equation, we find that the frequency of the molecule can be expressed by

$$\nu = \frac{Kc}{2\left(\dfrac{Mv_0}{N}\right)^{1/3} \times \dfrac{v-v_0}{v_0}} = \frac{K'c}{v-v_0}.$$

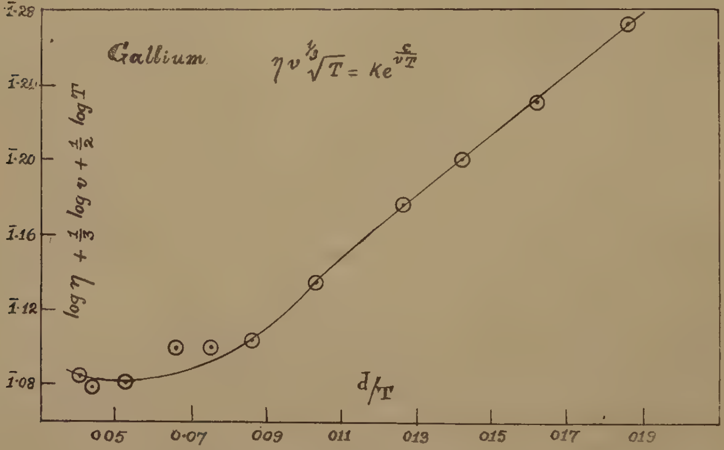
Again, since c varies directly as the square root of absolute temperature, the equation can be written in the form $\nu = \dfrac{K''\sqrt{T}}{v-v_0}$. Substituting this equation in Andrade's equation $\eta = \frac{4vm}{3\sigma}$, and putting $\sigma = Kv^{1/3}$, the equation giving the variation of viscosity with temperature assumes the form

$$\frac{\eta v^{1/3}}{\sqrt{T}} = \frac{B}{v - v_0},$$

where B is a constant. This equation gives good results when applied to fused metals, but not in the case of unassociated organic liquids ⁽²²⁾.

Liquid.	Diethyl ketone.	Octane.	Carbon tetrachloride.
Max. % diff.....	1·6	7·0	3·5

Fig. 2.



It is very doubtful that the differences are due to molecular association. Perhaps they are due to molecules not being even approximately spherical and to the increase in the size of molecules with temperature.

If, however, $v - v_0$ is assumed to be the same as $v - b$ in van der Waals' equation, then $\frac{1}{v - v_0} = \frac{M\pi_0 e^{c'/Tv}}{RT}$, as shown by Macleod ⁽²²⁾. Substituting this value of $v - b$ in the equation $\frac{\eta v^{1/3}}{\sqrt{T}} = \frac{B}{v - v_0}$, the following equation is obtained: $\eta v^{1/3}\sqrt{T} = Ke^{c/Tv}$. Even if the molecular weight is not constant but varies slightly according to the equation $M = M_0 e^{c/Tv}$, the form of the above equation will remain the same. This equation, while it represents organic liquids well ⁽²²⁾, fails in the case of liquid gallium, as fig. 2 shows. Obviously there is something wrong either with Macleod's conception of the variation of ν or his equating $v-v_0$ with $v-b$, or both. Over short ranges of temperature, as in organic liquids, the mistake does not become apparent.

Another equation that deserves consideration is that of Madge (23). Starting with the idea that the viscosity of liquids cannot be regarded as being due to diffusion of molecules, as in the case of gases, but as the limiting case of an elastic solid when the material breaks down under a stress, Madge derives the equation $\eta=\frac{Ae^{\beta T}}{T-b}$ after making a number of assumptions. This equation also represents the variation of viscosity very well, but only in cases where Andrade's simple equation holds. Madge himself has calculated the viscosity of ethyl alcohol and water (over short temperature ranges), octane, propylene bromide and benzene. The results agree well with experimental values. We have further tested the equation with some more liquids for which Andrade's equations have been tried (*vide* Table I.). The results are summarized below in three cases :—

Liquid.	η range.	Temp. range ° C.	A.	b.	$-\beta\times10^5$.	Max. % diff.
Ethylene chloride	2·6	0– 80	1·335	206	207	0·5
Carbontetrachloride.......	2·5	0– 70	1·750	210	263	0·3
Isobutylene bromide	6·4	0–140	3·365	230	308	1·8 *

The constant b was found by the formula

$$b=T_1+a-a\sqrt{\frac{\eta_1\eta_3}{\eta_1\eta_3-\eta_2^2}},$$

where η_1, η_2 and η_3 are viscosities corresponding to T_1, T_2 and T_3, each separated from the next by $a°$. Knowing the value of b, the constants A and β are found graphically by plotting log η (T$-b$) against T.

The equation fails in the case of associated liquids :

Liquid.	Temp. range ° C.	Max. % diff.
Water (10)	$-$ 9·3 to 100	5·0
Ethyl alcohol	$-$98·11 to 73·57	14·8
Propyl alcohol	$-$60·23 to 95·59	17·3

* This error is only in one case. In other cases it is below 0·9.

If a satisfactory agreement in the case of unassociated liquids is the criterion, Madge's equation is as good as Andrade's. However, when the equation is applied to fused metals (Table II.) it is found that though the agreement between the observed and calculated results is not worse than with Andrade's equations, the value of β becomes positive, thereby indicating that the rigidity modulus goes on increasing with temperature, which seems absurd.

Recently Irany (24) has approached the problem of temperature variation of viscosity of liquids in a novel way. He assumes that in the case of unassociated liquids a certain function of viscosity ϕ is linear in temperature, and its rate of variation with temperature is the same for all simple liquids. The function has not been mathematically defined but graphically worked out. A straight line is *arbitrarily* drawn on a squared paper and is assumed to represent the linear variation of ϕ with temperature. An unassociated liquid of low viscosity is taken and the temperature is linearly marked along the t axis. The points on the ϕ-axis corresponding to the various temperatures are made to represent the viscosity of the liquid in centipoises on the ϕ-scale. Thus ϕ is marked out corresponding to the different values of η in centipoises. Taking a common viscosity value of a second liquid having a higher but overlapping range of viscosity, and marking the corresponding points on the t-axis as the temperature of the second liquid, the scale is extended. The ϕ-function works out very well for organic liquids but breaks down when applied to fused metals. Fig. 3 shows the plot of ϕ against t for fused gallium and tin.

Fig. 3.

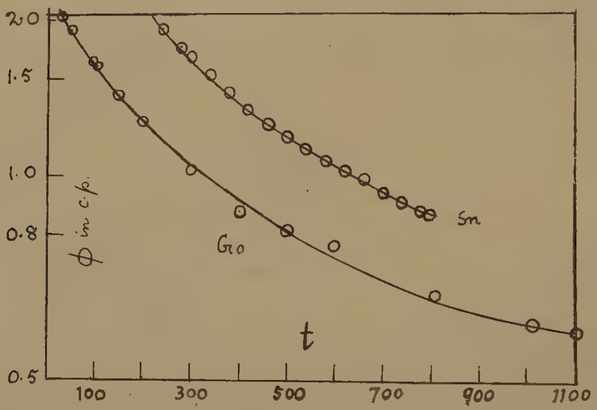


Table II. compares the results obtained with various equations for fused sodium, potassium, gallium and tin. Chiong's data (25) for sodium and potassium, Spells' data for gallium (26), and Stott's data for tin (27) have been used. For specific volumes the relative values given in Macleod's paper (22) are used in all cases except gallium, for which density data are available in Spells' paper. Mercury has not been tested due to uncertainty of its values. In Table III. detailed results for gallium are given. In both the Tables. I. refers to Andrade's simple formula, II. Andrade's modified formula. III, Madge's formula, IV. Silverman's simple formula and V. Silverman's modified formula. The percentage difference is given as $\frac{\eta_{\text{cal.}}-\eta_{\text{obs.}}}{\eta_{\text{obs.}}}\times 100$.

Table II.

	Temp. range. °K.	Sodium 371–628.	Potassium 337–625.	Gallium 303–1373.	Tin 513–1073.
I.	$A\times10^5$	92·53	99·20	415	434·5
	b	748·3	566·4	506·6	767·4
	Max. % diff.	1·0	0·8	4·1	2·1
* II.	$A\times10^3$	1·183	1·293	2·468	4·676
	c	716·5	600	79·5	732·2
	Max. % diff.	0·7	0·6	3·9	2·1
III.	A	1·021	0·71121	4·066	5·550
	b	203	166	75	174
	$\beta\times10^4$	3·643	7·391	4·462	3·198
	Max. % diff.	0·83	0·77	3·8	1·1
IV.	$K\times10^2$	2·808	3·042	95·41	17·64
	Q/R	579·1	394·1	253·3	461 6
	Max. % diff.	0·8	1·1	3·7	0·7
V.	$K\times10^2$	2·763	2·710	95·06	17·98
	Q/R	576·8	423·2	247·45	454·3
	$c\times10^6$	−43·91	−104·1	−4·62	9·74
	Max. % diff.	0·9	0·8	2·8	0·7

The first three values for sodium, the first for potassium and the first for gallium (in the papers already referred to) refer to temperatures within five degrees of their melting points. If these are included the percentage error increases a good deal as shown below.

	I.	II.	III.	IV.	V.
Sodium	4·2	4·4	4·2	4·6	5·0
Potassium	3·8	3·6	3·9	4·7	8·0
Gallium	8·4	8·5	3·8	3·7	2·8

---

* The results for modified Andrade's equation, except in the case of tin are reproduced from the original papers from which data have been taken.

TABLE III.

Gallium.		$\eta_{cal.} \times 10^3$.					% difference.				
Temp. °K.	$\eta_0 \times 10^3$.	I.	II.	III.	IV.	V.	I.	II.	III.	IV.	V.
303	*20·37	22·08	22·11	20·42	20·75	20·31	8·4	8·5	0·3	1·9	0·3
325·9	18·94	19·64	19·68	18·75	18·87	18·53	3·7	3·9	1·0	0·4	2·2
375	16·04	16·02	16·06	16·02	16·03	15·75	0·1	0·1	0·1	0·1	1·8
422	14·06	13·78	13·82	14·15	14·09	13·88	2·0	1·7	0·6	0·2	1·3
473	12·58	12·11	12·15	12·62	12·54	12·37	3·7	3·4	0·3	0·3	1·6
574	10·29	10·03	11·06	10·53	10·48	10·37	2·6	2·2	2·3	1·9	0·8
675	8·820	8·789	8·807	9·158	9·150	9·068	0·4	0·2	3·8	3·7	2·8
773	8·127	7·991	7·997	8·224	8·239	8·177	1·7	1·6	1·2	1·4	0·6
873	7·637	7·413	7·409	7·523	7·548	7·498	2·9	3·0	1·5	1·2	1·8
1079	6·524	6·636	6·608	6·557	6·554	6·527	1·7	1·3	0·5	0·5	0·0
1283	5·915	6·158	6·107	5·969	5·912	5·895	4·1	3·3	0·9	0·1	0·3
1373	5·783	6·001	5·940	5·782	5·689	5·678	3·7	2·7	0·0	1·6	1·7

* Printed by mistake as 20·27 in Macleod's paper (22).

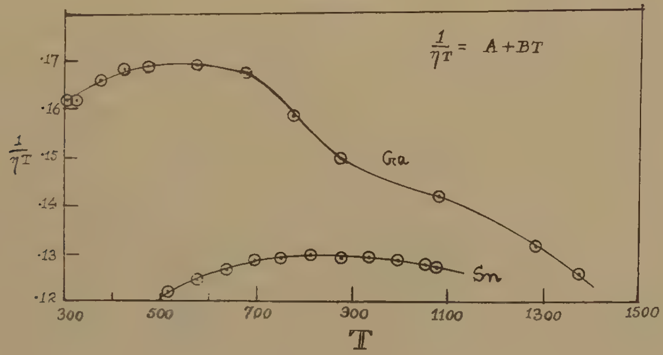
Many other equations not so successful as those discussed above have been proposed. Of these Frenkel's equation (28), which can be written as $\eta = \mathrm{AT}e^{b/\mathrm{T}}$, works all right when applied to liquids over short ranges. It represents octane in the form $\eta = 1{\cdot}607 \times 10^{-7}\mathrm{T}e\frac{1392}{\mathrm{T}}$ to an accuracy of 1·9 per cent. When applied to metals such as gallium it fails.

The equation of Mott Sauders (Jr.) (29); $\log \log \eta = md - 2.9$, where η is referred in millipoises, does not apply accurately even to unassociated organic liquids for the entire range over which measurements have been made.

Newton Friend (30) has proposed an equation $\frac{1}{\eta} = \mathrm{K}(v^{8/3} - v_0{}^{8/3})$, where v is specific volume. v_0 is a constant which is theoretically the specific volume when the viscosity is infinite. His equation also fails in the case of metals with the exception of tin (30).

Bingham and Stookey (31) have recently proposed $\frac{1}{\eta\mathrm{T}} = \mathrm{A} + \mathrm{BT}$, which represents simple organic liquids well but does not apply to fused metals, as the following figure shows.

Fig. 4.



Conclusions.

It is evident from what has been said that of all the equations proposed so far only those of Andrade, Silverman and Madge are applicable to unassociated organic and inorganic liquids over extended ranges of temperature.

Of these Andrade's simple and modified equations apply with almost the same accuracy. Silverman's modified equation leads to anomalous conclusions at times regarding compressibility. Thus with sodium, potassium and gallium it indicates that compressibility decreases with temperature, though it represents the results well. Silverman's simple equation gives as good results as the modified equation, and so is to be preferred to the modified equation. Madge's equation too gives very good results. but since it has to be assumed that the rigidity modulus increases with temperature at least in the case of metals, it has to be discarded on theoretical grounds. The simple equation of Andrade and the simple equation of Silverman give equally accurate results. If the former has to be preferred it is because it does not require any knowledge of density. Of all the equations that represent the variation of viscosity with temperature, there is no doubt that Andrade's simple formula is the simplest. Also it is not less accurate than any other formula.

So far as the fused metals are concerned Macleod's equation developed from Andrade's ideas of viscosity at the melting point, which is in a form similar to Batschinski's equation, deserves mention. In a way it is another vindication of Andrade's conception of viscosity.

One of us (M. K. S.) wishes to thank the Government of Orissa for the grant of a scholarship.

Summary.

1. A number of equations connecting viscosity and temperature have been examined.

2. It is found that Andrade's simple equation $\eta=Ae^{b/T}$ is as accurate as Andrade's modified equation $\eta v^{1/3}=Ae^{cv/T}$.

3. Of all the equations of Macleod only $\frac{\eta v^{1/3}}{\sqrt{T}}=\frac{B}{v-v_0}$ holds for fused metals. This equation, which holds for wide ranges of temperature for metals, fails in the case of organic liquids, probably due to the departure of organic molecules from spherical shape and the increase in size of the molecules with temperature.

4. Silverman's simple equation, $\eta=\frac{K'v}{T}e^{Q/RT}$. requiring a knowledge of density, gives good results. His modified equation, though equally applicable, leads to the erroneous conclusion that compressibility decreases with temperature in many cases.

5. Madge's equation, $\eta=\frac{A\epsilon^{\beta T}}{T-b}$, gives equally good results, but here again in the case of fused metals examined the modulus of rigidity has to be assumed to increase with temperature. Hence it is theoretically unsound.

6. Irany's function of viscosity is not linear in the case of fused metals.

Taking all these into account, Andrade's simple equation has to be considered best for unassociated liquids. It has the support of theory and it gives results as accurate as, if not better than, any other successful equation.

None of these equations can be applied to associated liquids, whose association changes rapidly with temperature.

References.

(1) Hatschek, 'The Viscosity of Liquids,' Bell.
(2) Ferguson, 'Reports on Progress in Physics,' ii. (1935).
(3) 'First Report on Viscosity and Plasticity,' Academy of Sciences, Amsterdam (1935).
(4) 'Second Report on Viscosity and Plasticity,' *ibid.* (1938).
(5) Porter, Phil. Mag. xxiii. p. 458 (1912).
(6) Guzman, *Anales de la Sociedad Espanola de Fisica y Quimica*, xi. p. 353 (1913).
(7) Kendall and Munroe, J. Amer. Chem. Soc. xxxix. p. 1799 (1917).
(8) Raman, Nature, cxi. p. 532 (1923).
(9) Dunn, Trans. Faraday Soc. xxii. p. 401 (1926).
(10) Andrade, Nature, (March 1, 1930) ; *ibid.* (April 12, 1930).
(11) Sheppard, Nature, (March 29, 1930).
(12) Andrade, Phil. Mag. (7), xvii. p. 497 (1934) ; (7), xvii. p. 698 (1934).
(13) Philips, Proc. Roy. Soc. A, lxxxvii. p. 48 (1912) ; Michels and Gibson, *ibid.* A, cxxxiv. p. 288 (1931).
(14) Prasad, J. Ind. Chem. Soc. x. p. 143 (1933) ; M. P. V. Iyer, Ind. J. Physics, v. p. 371 (1930).
(15) Prasad, J. Ind. Chem. Soc. xi. p. 913 (1934).
(16) Eyring, J. Chem. Phys. iv. p. 283 (1936).
(17) Lewis, J. Chem. Soc. p. 1061 (1938).
(18) Silverman, Trans. Faraday Soc. xxix. p. 1285 (1933).
(19) Macleod, Trans. Faraday Soc. xix. p. 6 (1923).
(20) Macleod, Trans. Faraday Soc. xxi. p. 151 (1925).
(21) Macleod, Trans. Faraday Soc. xxxii. p. 872 (1936).
(22) Macleod, Proc. Phys. Soc. l. p. 788 (1938).
(23) Madge, J. Physic. Chem. xxxiv. p. 1599 (1930).
(24) Irany, J. Amer. Chem. Soc. lx. p. 106 (1938).
(25) Chiong, Proc. Roy. Soc. A, clvii. p. 264 (1936).
(26) Spells, Proc. Phys. Soc. xlviii. p. 299 (1936).
(27) Stott, Proc. Phys. Soc. xlv. p. 530 (1933).
(28) Frenkel, *Zeit. f. Phys.* xxxv. p. 664 (1936).
(29) Mott Sauders (Jr.), J. Amer. Chem. Soc. lx. p. 154 (1938).
(30) Newton Friend, Trans. Faraday Soc. xxxiv. p. 813 (1938).
(31) Bingham and Stookey, J. Amer. Chem. Soc. lxi. p. 1625 (1939).

Mayurbunj Chemical Laboratory,
Ravenshaw College,
Cuttack, India.

XXVIII. *Non-Linear Distortion, with Particular Reference to the Theory of Frequency Modulated Waves.*—Part II.

By E. C. Cherry, M.Sc., and R. S. Rivlin, M.A.

(Communication from the Staff of the Research Laboratories of
The General Electric Company, Limited, England *.)

[Received December 8, 1941.]

Summary.

The output from a four-terminal network, having linear or non-linear amplitude-frequency and phase-shift-frequency characteristics, when a frequency modulated wave of E.M.F. is applied to the input, is discussed on a steady state theory. The results are not limited to the condition of carrier frequency large compared with frequency excursion. It is shown that the output E.M.F. is, in general, a " hybrid " wave in which both the frequency modulation and amplitude modulation terms are distorted. However, for the case of a linear phase-shift-frequency characteristic the output E.M.F. is an undistorted frequency modulated wave for all ratios of modulating frequency to carrier. For a linear amplitude-frequency characteristic the output E.M.F. is a " hybrid " wave in which both the frequency and amplitude modulation terms are substantially undistorted provided the carrier frequency is large compared with the frequency excursion.

1. *Introduction.*

In Part I. of this paper the problem was considered of the application of a frequency modulated wave of E.M.F. to a device such as a thermionic valve having a non-linear resistive characteristic. The purely sinusoidal components of the spectrum of a carrier wave, modulated in frequency in any arbitrary manner, were determined, and calculations were made of the components of the spectrum at the output terminals of the non-linear device when such a wave was applied to the input terminals. It was shown how the results were relevant to the theory of valve amplifiers, frequency multipliers and limiters such as are used in communication systems employing frequency modulation.

---

* Communicated by C. C. Paterson, F.R.S.

We are now concerned, in this part of the paper, with the application of a carrier wave, again modulated in any desired way, to devices having transmission characteristics which are non-linear functions of frequency, such as band-pass filters and converters (for converting frequency modulation into equivalent amplitude modulation).

This problem has been discussed by Roder (2), using a vectorial method, for the special case of a frequency modulated wave in which the modulating signal consists of a single sinusoidal component, and for particular forms of transmission network.

In the present paper we first deal with the case of an ideal converter (in Sections 2, 3, and 4). In the following sections we deal with the effects on the frequency modulated wave of non-linearity in the transmission characteristics. This analysis is not limited to the condition that the carrier frequency be large compared with the modulating frequencies. The manner in which difficulties arising from the consideration of " negative " frequencies may be avoided is discussed in Section 6.

Two methods of expressing the non-linear amplitude-frequency characteristic are used—a power series and a Fourier series ; the characteristic may be represented over any desired range of frequencies, and with any degree of accuracy, by taking a sufficient number of terms in either of these series. We use only a Fourier series representation of the non-linear phase-shift-frequency characteristic and restrict ourselves to small departures from linearity in this characteristic.

The general problem of the application of a frequency modulated wave, in which the modulating wave has any wave-form, to a network having a non-linear complex admittance-frequency characteristic has been discussed by Carson and Fry (3). They have developed a variable frequency circuit theory and, applying it to this problem, show that the current wave through the non-linear complex admittance is a " hybrid " wave consisting of the undistorted frequency-modulated wave, modulated in amplitude by the distorted modulating signal. This result disagrees with those we obtain in Section 5 (*b*) and 5 (*d*), where we consider a non-linear amplitude-frequency characteristic which can be represented either by a finite number of terms of a Fourier series or a polynomial. In these cases we have shown that the frequency modulated term in the " hybrid " wave is also distorted.

2. *The Application of a Frequency Modulated Wave to a Four-Terminal Network having a Linear Amplitude-Frequency Characteristic.*

Consider a frequency modulated wave of E.M.F. given by

$$v=\mathrm{V}\sin\left[\omega_0 t+\sum_{s=1}^{\mathrm{S}} m_s\cos(\omega_s t+\epsilon_s)\right] \quad . \quad . \quad . \quad . \quad . \quad (1)$$

applied to the input terminals of a four-terminal network having the linear amplitude-frequency characteristic given by

$$v_0=v_i(c_0+c_1\omega), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where $v_0 \sin \omega t$ is the output E.M.F. resulting from the application of the E.M.F. $v_i \sin \omega t$ to the input terminals of the network. Assume that the four-terminal network introduces zero phase-shift at all frequencies.

We have seen in Section 4 of Part I. of this paper, that the frequency modulated wave of equation (1) can be represented as a spectrum of purely sinusoidal E.M.F.'s, thus

$$v=\sum_{\{\alpha_s\}} \mathrm{V}\,.\left[\prod_{s=1}^{\mathrm{S}} \mathrm{J}_{\alpha_s}(m_s)\right].\cos\left(\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{s=1}^{\mathrm{S}}\alpha_s\phi_s\right) \quad . \quad . \quad (3)$$

with the notation $\phi_s=\omega_s t+\epsilon_s$, the summation $\sum\limits_{\{\alpha_s\}}$ being carried out over all possible sets of values of $\{\alpha_s\}$ where α_s may have any positive or negative integral value, or the value zero.

Now, if this wave is applied to the four-terminal network whose amplitude-frequency characteristic is given in equation (2), the output E.M.F. is

$$v_0=\sum_{\{\alpha_s\}} \mathrm{V}\,.\left[\prod_{s=1}^{\mathrm{S}} \mathrm{J}_{\alpha_s}(m_s)\right].\left[c_0+c_1\left(\omega_0+\sum_{s=1}^{\mathrm{S}}\alpha_s\omega_s\right)\right]$$
$$\times\cos\left(\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{s=1}^{\mathrm{S}}\alpha_s\phi_s\right). \quad (4)$$

From Appendix 3, it may be seen that each of the S series

$$\sum_{\{\alpha_s\}} \mathrm{V}\,.\left[\prod_{s=1}^{\mathrm{S}} \mathrm{J}_{\alpha_s}(m_s)\right].\,c_1\alpha\,\omega_r\,.\cos\left(\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{=1}^{\mathrm{S}}\alpha_s\phi_s\right),$$

where r can take any of the values 1 to S, and the series

$$\sum_{\{\alpha_s\}} \mathrm{V}\,.\left[\prod_{s=1}^{\mathrm{S}} \mathrm{J}_{\alpha_s}(m_s)\right].\,(c_0+c_1\omega_0)\,.\cos\left(\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{s=1}^{\mathrm{S}}\alpha_s\phi_s\right)$$

is absolutely convergent. Hence, the series (4) is absolutely convergent. Thus, the series (4) may be re-arranged to give

$$v_0=(c_0+c_1\omega_0)\,.\sum_{\{\alpha_s\}} \mathrm{V}\,.\left[\prod_{s=1}^{\mathrm{S}} \mathrm{J}_{\alpha_s}(m_s)\right].\cos\left(\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{s=1}^{\mathrm{S}}\alpha_s\phi_s\right)$$
$$+c_1\,.\sum_{r=1}^{\mathrm{S}}\omega_r\,.\left\{\sum_{\{\alpha_s\}} \mathrm{V}\,.\left[\prod_{s=1}^{\mathrm{S}} \mathrm{J}_{\alpha_s}(m_s)\right].\,\alpha_r\,.\cos\left(\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{s=1}^{\mathrm{S}}\alpha_s\phi_s\right)\right\}.$$

Now,

$$(c_0+c_1\omega_0)\,.\sum_{\{\alpha_s\}} \mathrm{V}\,.\left[\prod_{s=1}^{\mathrm{S}} \mathrm{J}_{\alpha_s}(m_s)\right].\cos\left(\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{s=1}^{\mathrm{S}}\alpha_s\phi_s\right)$$

$$=V(c_0+c_1\omega_0)\,.\,\sin\left(\omega_0 t+\sum_{s=1}^{S} m_s\cos\phi_s\right).$$

Again,

$$\sum_{\{\alpha_s\}} V\,.\left[\prod_{s=1}^{S} J_{\alpha_s}(m_s)\right].\,\alpha_r\,.\,\cos\left(\omega_0 t+\overline{\alpha-1}\,.\,\frac{\pi}{2}+\sum_{s=1}^{S}\alpha_s\phi_s\right)$$

$$=\sum_{\{\alpha_s\}} V\,.\,\frac{\left[\prod_{s=1}^{S} J_{\alpha_s}(m_s)\right]}{J_{\alpha_r}(m_r)}\,.\,\tfrac{1}{2}m_r\,.\,[J_{\alpha_r+1}(m_r)+J_{\alpha_r-1}(m_r)]$$

$$\times\cos\left(\omega_0 t+\overline{\alpha-1}\,.\,\frac{\pi}{2}+\sum_{s=1}^{S}\alpha_s\phi_s\right),$$

which may be rewritten as the sum of two absolutely convergent series; thus

$$\tfrac{1}{2}m_r\sum_{\{\alpha_s\}} V\,.\,\frac{\left[\prod_{s=1}^{S} J_{\alpha_s}(m_s)\right]}{J_{\alpha_r}(m_r)}\,.\,J_{\alpha_r+1}(m_r)\,.\,\cos\left[\omega_0 t+\alpha\,.\,\frac{\pi}{2}\right.$$

$$\left.-\phi_r-\frac{\pi}{2}+\left(\sum_{s=1}^{S}\alpha_s\phi_s+\phi_r\right)\right]$$

$$+\tfrac{1}{2}m_r\,.\sum_{\{\alpha_s\}} V\,.\,\frac{\left[\prod_{s=1}^{S} J_{\alpha_s}(m_s)\right]}{J_{\alpha_r}(m_r)}\,.\,J_{\alpha_r-1}(m_r)\,.\,\cos\left[\omega_0 t+\overline{\alpha-2}\,.\,\frac{\pi}{2}\right.$$

$$\left.+\phi_r+\frac{\pi}{2}+\left(\sum_{s=1}^{S}\alpha_s\phi_s-\phi_r\right)\right]$$

$$=\tfrac{1}{2}m_r\,.\,V\,.\left[\sin\left(\omega_0 t-\phi_r-\frac{\pi}{2}+\sum_{s=1}^{S} m_s\cos\phi_s\right)\right.$$

$$\left.+\sin\left(\omega_0 t+\phi_r+\frac{\pi}{2}+\sum_{s=1}^{S} m_s\cos\phi_s\right)\right]$$

$$=V\,.\,m_r\,.\,\cos\left(\phi_r+\frac{\pi}{2}\right).\,\sin\left(\omega_0 t+\sum_{s=1}^{S} m_s\cos\phi_s\right).$$

We thus see that the output

$$v_0=V\left[(c_0+c_1\omega_0)-c_1\sum_{s=1}^{S} m_s\omega_s\sin\phi_s\right].\,\sin\left(\omega_0 t+\sum_{s=1}^{S} m_s\cos\phi_s\right). \qquad (5)$$

3. *The Application of a Frequency Modulated Wave to a Four-Terminal Network having a Linear Phase-Shift-Frequency Characteristic.*

The frequency modulated wave of equation (1) is applied to the input terminals of a four-terminal network having the linear phase-shift frequency characteristic given by

$$\psi_0=\eta_0+\eta_1\omega, \quad . \; . \; . \; . \; . \; . \; . \; . \; . \qquad (6)$$

and zero attenuation at all frequencies, so that the output E.M.F. resulting from the application of the E.M.F. $v_i \cos \omega t$ to the input terminals, is $v_i \cos (\omega t+\psi_0)$. In any physical network η_0 will be zero or an integral multiple of $\pi/2$.

Using the expression (3) for the frequency modulated wave, we see that the output E.M.F. may be written

$$\begin{aligned} v_0 &= \sum_{\{\alpha_s\}} \mathrm{V} \,.\left[\prod_{s=1}^{\mathrm{S}} \mathrm{J}_{\alpha_s}(m_s)\right].\cos\left[\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{s=1}^{\mathrm{S}}\alpha_s\phi_s\right. \\ &\qquad\qquad \left.+\eta_0+\eta_1\left(\omega_0+\sum_{s=1}^{\mathrm{S}}\alpha_s\omega_s\right)\right], \\ &= \sum_{\{\alpha_s\}} \mathrm{V} \,.\left[\prod_{s=1}^{\mathrm{S}} \mathrm{J}_{\alpha_s}(m_s)\right].\cos\left[\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}+\eta_0+\eta_1\omega\right. \\ &\qquad\qquad \left.+\sum_{s=1}^{\mathrm{S}}\alpha_s(\phi_s+\eta_1\omega_s)\right], \\ &= \mathrm{V}\,.\sin\left[\omega_0 t+\eta_0+\eta_1\omega_0+\sum_{s=1}^{\mathrm{S}} m_s\cos(\phi_s+\eta_1\omega_s)\right]. \end{aligned}$$

This is seen to be the applied wave delayed by a time η_1, but undistorted, the unmodulated carrier being shifted in phase by $\eta_0+\eta_1\omega_0$.

4. *The Application of a Frequency Modulated Wave to a Four-Terminal Network having Linear Amplitude-Frequency and Phase-Shift-Frequency Characteristics.*

Suppose the frequency modulated wave of equation (1) is applied to the input terminals of a four-terminal network having the linear amplitude-frequency characteristic of equation (2) and the linear phase-shift-frequency characteristic of equation (6).

We can consider the network to consist of two networks in tandem. the first having the linear phase-shift-frequency characteristic of equation (6) and introducing zero attenuation for all frequencies, the second having the linear amplitude-frequency characteristic of equation (2) and introducing zero phase-shift at all frequencies. The combination will then have the given amplitude and phase-shift characteristics.

From Section 3, we see that if the frequency modulated wave of equation (1) is applied to a four-terminal network having the phase-shift characteristic of equation (6), the output E.M.F. is

$$\mathrm{V}\sin\left[\omega_0 t+\eta_0+\eta_1\omega_0+\sum_{s=1}^{\mathrm{S}} m_s\cos(\phi_s+\eta_1\omega_s)\right].$$

When this is applied to a four-terminal network having the amplitude characteristic of equation (2), the output E.M.F. is obtained from equation (5), by substituting $(\omega_0 t+\eta_0+\eta_1\omega_0)$ for $\omega_0 t$ and $(\phi_s+\eta_1\omega_s)$ for ϕ_s, giving

$$\begin{aligned}
&(c_0+c_1\omega_0)\,.\,\mathrm{V}\,.\,\sin\left[\omega_0 t+\eta_0+\eta_1\omega_0+\sum_{s=1}^{S} m_s\cos(\phi_s+\eta_1\omega_s)\right]\\
&\quad -c_1\,.\,\mathrm{V}\,.\left[\sum_{s=1}^{S} m_s\omega_s\,.\,\sin(\phi_s+\eta_1\omega_s)\right].\,\sin\left[\omega_0 t+\eta_0+\eta_1\omega_0\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\left.+\sum_{s=1}^{S} m_s\cos(\phi_s+\eta_1\omega_s)\right],\\
&=(c_0+c_1\omega_0)\,.\,\mathrm{V}\,.\,\sin\left[\omega_0 t+\eta_0+\eta_1\omega_0+\sum_{s=1}^{S} m_s\cos(\phi_s+\eta_1\omega_s)\right]\\
&\quad +c_1\omega_0\mathrm{V}\left[\sum_{s=1}^{S} k_s\sin(\phi_s+\eta_1\omega_s)\right].\,\sin\left[\omega_0 t+\eta_0+\eta_1\omega_0\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\left.+\sum_{s=1}^{S} m_s\cos(\phi_s+\eta_1\omega_s)\right].
\end{aligned}$$

This consists of two components :

$$(1)\quad (c_0+c_1\omega_0)\,.\,\mathrm{V}\,.\,\sin\left[\omega_0 t+\eta_0+\eta_1\omega_0+\sum_{s=1}^{S} m_s\cos(\phi_s+\eta_1\omega_s)\right],$$

which is the applied frequency modulated wave, modified in amplitude by the factor $(c_0+c_1\omega_0)$ and delayed as shown in Section 3.

$$\begin{aligned}
(2)\quad & c_1\omega_0\mathrm{V}\left[\sum_{s=1}^{S} k_s\sin(\phi_s+\eta_1\omega_s)\right].\,\sin\left[\omega_0 t+\eta_0\right.\\
&\qquad\qquad\qquad\qquad\left.+\eta_1\omega_0+\sum_{s=1}^{S} m_s\cos(\phi_s+\eta_1\omega_s)\right],
\end{aligned}$$

which is a " hybrid " wave consisting of the applied frequency modulated wave (delayed in time) modulated in *amplitude* by the modulating function

$$\sum_{s=1}^{S} k_s\sin(\phi_s+\eta_1\omega_s),$$

which is the original modulating function delayed by time η_1.

If this output E.M.F. is applied to a detector, the original modulating signal may be obtained. The only distortion introduced will be that produced by the detector.

5. *The Application of a Frequency Modulated Wave to a Four-Terminal Network having Non-Linear Amplitude-Frequency and Phase-Shift-Frequency Characteristics.*

(a) *Form of Transmission Characteristics.*

The transmission properties of a four-terminal network can, in general, be completely defined in terms of its amplitude-frequency and phase-shift-frequency characteristics. In Section 4 we have discussed the case where both of these characteristics can be represented by linear functions of frequency. Such a network provides an ideal discriminator for frequency modulated waves. If these characteristics are no longer linear, the network will no longer behave as a perfect discriminator, the departures from linearity producing distortion in the envelope of the output wave.

These non-linear amplitude characteristics may be expressed over some range to any desired degree of approximation by an equation of either of the forms :

$$v_0 = v_i \sum_{n=0}^{N} c_n \omega^n, \quad \ldots \ldots \ldots \quad (7)$$

or

$$v_0 = v_i \left[c_0 + c_1\omega + \sum_{n=1}^{N} \left(g_n \cos\frac{2\pi n\omega}{W} + h_n \sin\frac{2\pi n\omega}{W} \right) \right]. \quad \ldots \quad (8)$$

W is a chosen range of angular frequency which must be sufficiently large to include all sideband components in the output of appreciable amplitude. For purposes of analysis this range may be chosen as large as may be desired, but in practice its necessary extent may be found from a consideration of the " band-width " of the applied wave (this is discussed in Section 4 of Part I. of this paper) together with the form of the characteristic.

The phase-shift-frequency characteristic may similarly be expressed to any desired degree of approximation in the form

$$\psi_0 = \eta_0 + \eta_1\omega + \sum_{n=1}^{N} \left(\lambda_n \cos\frac{2\pi n\omega}{W} + \mu_n \sin\frac{2\pi n\omega}{W} \right). \quad \ldots \quad (9)$$

(b) *The Effect of Non-Linear Amplitude-Frequency Characteristic.* (*Method* I.)

In this section the four-terminal network, to which the frequency modulated wave of equation (1) is applied, is considered to have an amplitude-frequency characteristic of the form (8), and an identically zero phase-shift-frequency characteristic.

The output E.M.F. is then given by

$$\begin{aligned} v_0 = \sum_{\{\alpha_s\}} V \,.\, \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s) \right] . \Bigg[c_0 + c_1\left(\omega_0 + \sum_{s=1}^{S} \alpha_s\omega_s \right) \\ + \sum_{n=1}^{N} \left\{ g_n \cos\frac{2\pi n}{W}\left(\omega_0 + \sum_{s=1}^{S} \alpha_s\omega_s \right) + h_n \sin\frac{2\pi n}{W}\left(\omega_0 + \sum_{s=1}^{S} \alpha_s\omega_s \right) \right\} \Bigg] \\ \times \cos\left(\omega_0 t + \overline{\alpha - 1} \,.\, \frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s\phi_s \right). \quad (10) \end{aligned}$$

Now, we know from Section 1 that the series

$$\sum_{\{\alpha_s\}} V \cdot \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s)\right] \cdot \left[c_0 + c_1\left(\omega_0 + \sum_{s=1}^{S} \alpha_s\omega_s\right)\right]$$
$$\times \cos\left(\omega_0 t + \overline{\alpha-1} \cdot \frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s\phi_s\right)$$

is absolutely convergent. Also, each of the series

$$\sum_{\{\alpha_s\}} V \cdot \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s)\right] \cdot g_n \cos\frac{2\pi n}{W}\left(\omega_0 + \sum_{s=1}^{S} \alpha_s\omega_s\right)$$
$$\times \cos\left(\omega_0 t + \overline{\alpha-1} \cdot \frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s\phi_s\right)$$

and
$$\sum_{\{\alpha_s\}} V \cdot \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s)\right] \cdot h_n \sin\frac{2\pi n}{W}\left(\omega_0 + \sum_{s=1}^{S} \alpha_s\omega_s\right)$$
$$\times \cos\left(\omega_0 t + \overline{\alpha-1} \cdot \frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s\phi_s\right)$$

is absolutely convergent since

$$\left|\cos\frac{2\pi n}{W}\left(\omega_0 + \sum_{s=1}^{S} \alpha_s\omega_s\right)\right| \leqslant 1 \quad \text{and} \quad \left|\sin\frac{2\pi n}{W}\left(\omega_0 + \sum_{s=1}^{S} \alpha_s\omega_s\right)\right| \leqslant 1$$

and the series

$$\sum_{\{\alpha_s\}} V \cdot \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s)\right] \cdot \cos\left(\omega_0 t + \overline{\alpha-1} \cdot \frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s\omega_s\right)$$

is absolutely convergent.

Thence, we can rearrange series (10) as

$$\sum_{\{\alpha_s\}} V \cdot \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s)\right] \cdot \left[c_0 + c_1\left(\omega_0 + \sum_{s=1}^{S} \alpha_s\omega_s\right)\right]$$
$$\times \cos\left(\omega_0 t + \overline{\alpha-1} \cdot \frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s\phi_s\right)$$
$$+ \sum_{n=1}^{N} \sum_{\{\alpha_s\}} V g_n \cdot \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s)\right] \cdot \cos\frac{2\pi n}{W}\left(\omega_0 + \sum_{s=1}^{S} \alpha_s\omega_s\right)$$
$$\times \cos\left(\omega_0 t + \overline{\alpha-1} \cdot \frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s\phi_s\right)$$
$$+ \sum_{n=1}^{N} \sum_{\{\alpha_s\}} V h_n \cdot \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s)\right] \cdot \sin\frac{2\pi n}{W}\left(\omega_0 + \sum_{s=1}^{S} \alpha_s\omega_s\right)$$
$$\times \cos\left(\omega_0 t + \overline{\alpha-1} \cdot \frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s\phi_s\right). \qquad (11)$$

Now, consider

$$\sum_{\{\alpha_s\}} V g_n \cdot \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s)\right] \cdot \cos\frac{2\pi n}{W}\left(\omega_0 + \sum_{s=1}^{S} \alpha_s\omega_s\right)$$
$$\times \cos\left(\omega_0 t + \overline{\alpha-1} \cdot \frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s\phi_s\right)$$

$$=\tfrac{1}{2}\underset{\{\alpha_s\}}{\Sigma} \mathrm{V}g_n \cdot \left[\prod_{s=1}^{\mathrm{S}} \mathrm{J}_{\alpha_s}(m_s)\right] . \cos\left[\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}+\frac{2\pi n\omega_0}{\mathrm{W}} + \sum_{s=1}^{\mathrm{S}} \alpha_s\left(\phi_s+\frac{2\pi n}{\mathrm{W}}\cdot\omega_s\right)\right]$$

$$+\tfrac{1}{2}\underset{\{\alpha_s\}}{\Sigma} \mathrm{V}g_n \cdot \left[\prod_{s=1}^{\mathrm{S}} \mathrm{J}_{\alpha_s}(m_s)\right] . \cos\left[\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}-\frac{2\pi n\omega_0}{\mathrm{W}} + \sum_{s=1}^{\mathrm{S}} \alpha_s\left(\phi_s-\frac{2\pi n}{\mathrm{W}}\cdot\omega_s\right)\right]$$

$$=\tfrac{1}{2}g_n\mathrm{V} . \left[\sin\left\{\omega_0 t+\frac{2\pi n}{\mathrm{W}}\,\omega_0+\sum_{s=1}^{\mathrm{S}} m_s\cos\left(\phi_s+\frac{2\pi n}{\mathrm{W}}\cdot\omega_s\right)\right\} + \sin\left\{\omega_0 t-\frac{2\pi n}{\mathrm{W}}\omega_0+\sum_{s=1}^{\mathrm{S}} m_s\cos\left(\phi_s-\frac{2\pi n}{\mathrm{W}}\cdot\omega_s\right)\right\}\right]. \quad (12)$$

Similarly, it may be shown that

$$\underset{\{\alpha_s\}}{\Sigma} \mathrm{V}h_n \cdot \left[\prod_{s=1}^{\mathrm{S}} \mathrm{J}_{\alpha_s}(m_s)\right] . \sin\frac{2\pi n}{\mathrm{W}}\left(\omega_0+\sum_{s=1}^{\mathrm{S}}\alpha_s\omega_s\right) \times\cos\left(\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{s=1}^{\mathrm{S}}\alpha_s\phi_s\right)$$

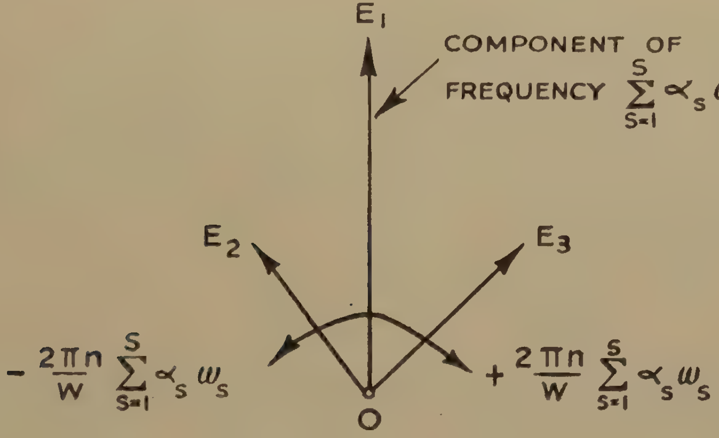
$$=-\tfrac{1}{2}h_n\mathrm{V} . \left[\cos\left\{\omega_0 t+\frac{2\pi n}{\mathrm{W}}\cdot\omega_0+\sum_{s=1}^{\mathrm{S}} m_s\cos\left(\phi_s+\frac{2\pi n}{\mathrm{W}}\cdot\omega_s\right)\right\} - \cos\left\{\omega_0 t-\frac{2\pi n}{\mathrm{W}}\cdot\omega_0+\sum_{s=1}^{\mathrm{S}} m_s\cos\left(\phi_s-\frac{2\pi n}{\mathrm{W}}\cdot\omega_s\right)\right\}\right]. \quad (13)$$

Thus, substituting the results of equations (11), (12), and (13) in equation (10), we have

$$\begin{aligned} v_0=(c_0+c_1\omega_0) . \mathrm{V} . \sin\left(\omega_0 t+\sum_{s=1}^{\mathrm{S}} m_s\cos\phi_s\right) &\\ -c_1\mathrm{V} . \left(\sum_{s=1}^{\mathrm{S}} m_s\omega_s\sin\phi_s\right) . \sin\left(\omega_0 t+\sum_{s=1}^{\mathrm{S}} m_s\cos\phi_s\right) &\\ +\sum_{n=1}^{\mathrm{N}}\left\{\tfrac{1}{2}g_n\mathrm{V} . \left[\sin\left\{\omega_0 t+\frac{2\pi n}{\mathrm{W}}\cdot\omega_0+\sum_{s=1}^{\mathrm{S}} m_s\cos\left(\phi_s+\frac{2\pi n}{\mathrm{W}}\cdot\omega_s\right)\right\}\right.\right. &\\ \left. +\sin\left\{\omega_0 t-\frac{2\pi n}{\mathrm{W}}\cdot\omega_0+\sum_{s=1}^{\mathrm{S}} m_s\cos\left(\phi_s-\frac{2\pi n}{\mathrm{W}}\cdot\omega_s\right)\right\}\right] &\\ -\tfrac{1}{2}h_n\mathrm{V} . \left[\cos\left\{\omega_0 t+\frac{2\pi n}{\mathrm{W}}\cdot\omega_0+\sum_{s=1}^{\mathrm{S}} m_s\cos\left(\phi_s+\frac{2\pi n}{\mathrm{W}}\cdot\omega_s\right)\right\}\right. &\\ \left.\left. -\cos\left\{\omega_0 t-\frac{2\pi n}{\mathrm{W}}\cdot\omega_0+\sum_{s=1}^{\mathrm{S}} m_s\cos\left(\phi_s-\frac{2\pi n}{\mathrm{W}}\cdot\omega_s\right)\right\}\right]\right\}. &\quad (14) \end{aligned}$$

We see that this consists of an undistorted frequency modulated wave and "hybrid" wave, identical with those obtained under the conditions of Section 2, together with a distortion term which consists of " echoes " of the applied wave delayed and advanced in time from the applied wave. Each component in the Fourier series (equation (8)) representing the non-linearity of the characteristic gives rise to one pair of " echoes," delayed and advanced by equal times, the time in each case being $\frac{2\pi n}{W}$.

This is illustrated in the figure ; OE_1 is a vector representing any one component of the applied frequency modulated wave. OE_2 and OE_3 are the two " echo " components which arise by virtue of one of the terms of the Fourier series (8). If OE_1 be the component of angular frequency $\sum_{s=1}^{S} \alpha_s \omega_s$, the " echoes " OE_2 and OE_3 (of the same frequency) will have


E_1
COMPONENT OF FREQUENCY $\sum_{S=1}^{S} \alpha_S \omega_S$
E_2
E_3
$-\frac{2\pi n}{W} \sum_{S=1}^{S} \alpha_S \omega_S$
$+\frac{2\pi n}{W} \sum_{S=1}^{S} \alpha_S \omega_S$
O

amplitudes $\frac{1}{2}g_n V$ or, $\frac{1}{2}h_n V$ and will be displaced from OE_1 by angles $\pm\frac{2\pi n}{W} \sum_{s=1}^{S} \alpha_s \omega_s$. These angles are, of course, proportional to the frequency of the component OE_1, the time interval between this component and each of the echoes being $\frac{2\pi n}{W}$.

The distortion due to each component of the Fourier series (8) may be expressed in an alternative form by summing the pair of " echoes " produced. Thus, the expression of equation (12)

$$\tfrac{1}{2}g_n V\left[\sin\left\{\omega_0 t+\frac{2\pi n}{W}\cdot\omega_0+\sum_{s=1}^{S} m_s \cos\left(\phi_s+\frac{2\pi n}{W}\omega_s\right)\right\}\right.$$
$$\left.+\sin\left\{\omega_0 t-\frac{2\pi n}{W}\cdot\omega_0+\sum_{s=1}^{S} m_s \cos\left(\phi_s-\frac{2\pi n}{W}\omega_s\right)\right\}\right]$$

$$=g_n\mathrm{V}\cos\left\{\frac{2\pi n}{\mathrm{W}}\cdot\omega_0-\sum_{s=1}^{\mathrm{S}}m_s\sin\frac{2\pi n}{\mathrm{W}}\omega_s\cdot\sin\phi_s\right\}$$
$$\times\sin\left\{\omega_0 t+\sum_{s=1}^{\mathrm{S}}m_s\cos\frac{2\pi n}{\mathrm{W}}\omega_s\cdot\cos\phi_s\right\}.\quad\quad(15)$$

Again, developing the expression of equation (13),

$$-\tfrac{1}{2}h_n\mathrm{V}\left[\cos\left\{\omega_0 t+\frac{2\pi n}{\mathrm{W}}\cdot\omega_0+\sum_{s=1}^{\mathrm{S}}m_s\cos\left(\phi_s+\frac{2\pi n}{\mathrm{W}}\omega_s\right)\right\}\right.$$
$$\left.-\cos\left\{\omega_0 t-\frac{2\pi n}{\mathrm{W}}\cdot\omega_0+\sum_{s=1}^{\mathrm{S}}m_s\cos\left(\phi_s-\frac{2\pi n}{\mathrm{W}}\omega_s\right)\right\}\right]$$
$$=h_n\mathrm{V}\cdot\sin\left\{\frac{2\pi n}{\mathrm{W}}\cdot\omega_0-\sum_{s=1}^{\mathrm{S}}m_s\sin\frac{2\pi n}{\mathrm{W}}\omega_s\cdot\sin\phi_s\right\}$$
$$\times\sin\left\{\omega_0 t+\sum_{s=1}^{\mathrm{S}}m_s\cos\frac{2\pi n}{\mathrm{W}}\omega_s\cdot\cos\phi_s\right\}.\quad\quad(16)$$

It thus appears that each component in the Fourier series (8) gives rise to a "hybrid" wave consisting of a frequency-modulated wave (which is different from the applied frequency-modulated wave in that the frequency modulated coefficient, for the modulating term of angular frequency ω_s, is modified from m_s to $m_s\cos\frac{2\pi n}{\mathrm{W}}\omega_s$) modulated in amplitude by a wave of the form

$$\begin{matrix}\cos\\ \sin\end{matrix}\left\{\frac{2\pi n}{\mathrm{W}}\omega_0-\sum_{s=1}^{\mathrm{S}}m_s\sin\frac{2\pi n}{\mathrm{W}}\omega_s\cdot\sin\phi_s\right\}.$$

This modulating wave contains components having the angular frequencies $\{\omega_s\}$ of the modulating signal together with harmonics and intermodulation products of these.

The output v_0 given in equation (14) may be expressed in the alternative form

$$\Sigma_1\cos\omega_0 t+\Sigma_2\sin\omega_0 t=\sqrt{\Sigma_1^{\,2}+\Sigma_2^{\,2}}\cdot\sin\left(\omega_0 t+\tan^{-1}\frac{\Sigma_1}{\Sigma_2}\right).\quad\quad(17)$$

where
$$\frac{\Sigma_1}{\mathrm{V}}=\left(c_0+c_1\omega_0-c_1\sum_{s=1}^{\mathrm{S}}m_s\omega_s\sin\phi_s\right)\cdot\sin\left(\sum_{s=1}^{\mathrm{S}}m_s\cos\phi_s\right)$$
$$+\sum_{n=1}^{\mathrm{N}}\tfrac{1}{2}g_n\left\{\sin\left[\frac{2\pi n}{\mathrm{W}}\omega_0+\sum_{s=1}^{\mathrm{S}}m_s\cos\left(\phi_s+\frac{2\pi n}{\mathrm{W}}\omega_s\right)\right]\right.$$
$$\left.-\sin\left[\frac{2\pi n}{\mathrm{W}}\omega_0-\sum_{s=1}^{\mathrm{S}}m_s\cos\left(\phi_s-\frac{2\pi n}{\mathrm{W}}\omega_s\right)\right]\right\}$$
$$-\sum_{n=1}^{\mathrm{N}}\tfrac{1}{2}h_n\left\{\cos\left[\frac{2\pi n}{\mathrm{W}}\omega_0+\sum_{s=1}^{\mathrm{S}}m_s\cos\left(\phi_s+\frac{2\pi n}{\mathrm{W}}\omega_s\right)\right]\right.$$
$$\left.-\cos\left[\frac{2\pi n}{\mathrm{W}}\omega_0-\sum_{s=1}^{\mathrm{S}}m_s\cos\left(\phi_s-\frac{2\pi n}{\mathrm{W}}\omega_s\right)\right]\right\}$$

and

$$\begin{aligned}\frac{\Sigma_2}{\mathrm{V}}=&\left(c_0+c_1\omega_0-c_1\sum_{s=1}^{\mathrm{S}} m_s\omega_s \sin\phi_s\right).\cos\left(\sum_{s=1}^{\mathrm{S}} m_s\cos\phi_s\right)\\ &+\sum_{n=1}^{\mathrm{N}}\tfrac{1}{2}g_n\left\{\cos\left[\frac{2\pi n}{\mathrm{W}}\omega_0+\sum_{s=1}^{\mathrm{S}} m_s\cos\left(\phi_s+\frac{2\pi n}{\mathrm{W}}\omega_s\right)\right]\right.\\ &\qquad\left.+\cos\left[\frac{2\pi n}{\mathrm{W}}\omega_0-\sum_{s=1}^{\mathrm{S}} m_s\cos\left(\phi_s-\frac{2\pi n}{\mathrm{W}}\omega_s\right)\right]\right\}\\ &-\sum_{n=1}^{\mathrm{N}}\tfrac{1}{2}h_n\left\{-\sin\left[\frac{2\pi n}{\mathrm{W}}\omega_0+\sum_{s=1}^{\mathrm{S}} m_s\cos\left(\phi_s+\frac{2\pi n}{\mathrm{W}}\omega_s\right)\right]\right.\\ &\qquad\left.+\sin\left[\frac{2\pi n}{\mathrm{W}}\omega_0-\sum_{s=1}^{\mathrm{S}} m_s\cos\left(\phi_s-\frac{2\pi n}{\mathrm{W}}\omega_s\right)\right]\right\}.\end{aligned}$$

We see from this that the output may be considered to be a " hybrid " wave consisting of a carrier, modulated in frequency by a distorted version of the original modulating signal and in amplitude by a different distorted version of the original modulating signal.

It is noteworthy that if we had considered the problem from the " instantaneous frequency " point of view, it would have appeared that non-linearity in the characteristic gives rise to distortion of the envelope, but not of the " instantaneous frequency " of the frequency-modulated wave.

(c) *The Effect of Non-Linear Phase-Shift-Frequency Characteristic.* (*Method* I.)

We shall now consider the application of the frequency-modulated wave of equation (1) to a four-terminal network having a non-linear phase-shift-frequency characteristic given by equation (9), and which introduces zero attenuation at all frequencies.

The characteristic of equation (9) may be rewritten

$$\psi_0=\eta_0+\eta_1\omega+\sum_{n=1}^{\mathrm{N}}\left[\nu_n\cos\left(\frac{2\pi n}{\mathrm{W}}\omega+\chi_n\right)\right],$$

where $\nu_n{}^2=\lambda_n{}^2+\mu_n{}^2$ and $\chi_n=\tan^{-1}\left(-\frac{\mu_n}{\lambda_n}\right)$.

It has been seen in Section 3 that a four-terminal network of linear phase-shift-frequency characteristic introduces no distortion in the frequency-modulated wave, but merely a time displacement. We can, therefore, in estimating the distortion produced by a non-linear phase-shift-frequency characteristic neglect the terms $\eta_0+\eta_1\omega$ and consider the characteristic as

$$\psi_0=\sum_{n=1}^{\mathrm{N}}\left[\nu_n\cos\left(\frac{2\pi n}{\mathrm{W}}\omega+\chi_n\right)\right]. \qquad \text{(18}$$

Now, the output from such a four-terminal network when the frequency-modulated wave of equation (1) is applied to it, is given by

$$v_0=\mathrm{V}\,.\sum_{\{\alpha_s\}}\left[\prod_{s=1}^{\mathrm{S}}\mathrm{J}_{\alpha_s}(m_s)\right].\cos\left[\omega_0t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{s=1}^{\mathrm{S}}\alpha_s\phi_s\right.$$
$$\left.+\sum_{n=1}^{\mathrm{N}}\nu_n\cos\left\{\frac{2\pi n}{\mathrm{W}}\left(\omega_0+\sum_{s=1}^{\mathrm{S}}\alpha_s\omega_s\right)+\chi_n\right\}\right].$$

Restricting ourselves to small values of ν_n, *i. e.*, the phase-shift-frequency characteristic departs only slightly from non-linearity, we have

$$v_0=\mathrm{V}\,.\sum_{\{\alpha_s\}}\left[\prod_{s=1}^{\mathrm{S}}\mathrm{J}_{\alpha_s}(m_s)\right].\left\{\cos\left[\omega_0t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{s=1}^{\mathrm{S}}\alpha_s\phi_s\right]\right.$$
$$-\sum_{n=1}^{\mathrm{N}}\nu_n\,.\cos\left[\frac{2\pi n}{\mathrm{W}}\cdot\left(\omega_0+\sum_{s=1}^{\mathrm{S}}\alpha_s\omega_s\right)+\chi_n\right]$$
$$\left.\times\sin\left(\omega_0t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{s=1}^{\mathrm{S}}\alpha_s\phi_s\right)\right\}$$
$$=\mathrm{V}\,.\sin\left(\omega_0t+\sum_{s=1}^{\mathrm{S}}m_s\cos\phi_s\right)$$
$$-\tfrac{1}{2}\sum_{n=1}^{\mathrm{N}}\nu_n\left\{\sum_{\{\alpha_s\}}\left[\prod_{s=1}^{\mathrm{S}}\mathrm{J}_{\alpha_s}(m_s)\right].\left[\sin\left(\omega_0t+\overline{\alpha-1}\,.\frac{\pi}{2}+\frac{2\pi n}{\mathrm{W}}\omega_0\right.\right.\right.$$
$$\left.+\chi_n+\sum_{s=1}^{\mathrm{S}}\alpha_s\,.\,\overline{\phi_s+\frac{2\pi n}{\mathrm{W}}\cdot\omega_s}\right)+\sin\left(\omega_0t+\overline{\alpha-1}\,.\frac{\pi}{2}-\frac{2\pi n}{\mathrm{W}}\omega_0\right.$$
$$\left.\left.\left.-\chi_n+\sum_{s=1}^{\mathrm{S}}\alpha_s\,.\,\overline{\phi_s-\frac{2\pi n}{\mathrm{W}}\cdot\omega_s}\right)\right]\right\}$$
$$=\mathrm{V}\sin\left(\omega_0t+\sum_{s=1}^{\mathrm{S}}m_s\cos\phi_s\right)$$
$$+\sum_{n=1}^{\mathrm{N}}\tfrac{1}{2}\nu_n\left\{\cos\left(\omega_0t+\frac{2\pi n}{\mathrm{W}}\omega_0+\chi_n+\sum_{s=1}^{\mathrm{S}}m_s\cos\overline{\phi_s+\frac{2\pi n}{\mathrm{W}}\omega_s}\right)\right.$$
$$\left.+\cos\left(\omega_0t-\frac{2\pi n}{\mathrm{W}}\omega_0-\chi_n+\sum_{s=1}^{\mathrm{S}}m_s\cos\overline{\phi_s-\frac{2\pi n}{\mathrm{W}}\cdot\omega_s}\right)\right\}.$$

This consists of an undistorted frequency modulated component, together with a pair of "echoes" arising from each term of the phase-shift-frequency characteristic (18).

The expression for v_0 may be rewritten as

$$\mathrm{V}\sin\left(\omega_0t+\sum_{s=1}^{\mathrm{S}}m_s\cos\phi_s\right)$$
$$+\sum_{n=1}^{\mathrm{N}}\nu_n\,.\cos\left(\frac{2\pi n}{\mathrm{W}}\omega_0+\chi_n-\sum_{s=1}^{\mathrm{S}}m_s\sin\frac{2\pi n}{\mathrm{W}}\omega_s\,.\sin\phi_s\right)$$
$$\times\cos\left(\omega_0t+\sum_{s=1}^{\mathrm{S}}m_s\cos\frac{2\pi n}{\mathrm{W}}\omega_s\,.\cos\phi_s\right).$$

(d) *The Effect of Non-Linear Amplitude-Frequency Characteristic* (*Method* II.).

The output obtained when the frequency modulated wave of equation (1) is applied to a four-terminal network of non-linear amplitude-frequency characteristic and identically zero phase-shift can be obtained by an alternative method to that given in Section 5 (b). The amplitude-frequency characteristic is now expressed in the form given in equation (7).

The output E.M.F. is

$$\begin{aligned} v_0 &= \mathrm{V}\,.\sum_{\{\alpha_s\}}\left[\prod_{s=1}^{\mathrm{S}}\mathrm{J}_{\alpha_s}(m_s)\right].\left[\sum_{n=0}^{\mathrm{N}}c_n\left(\omega_0+\sum_{s=1}^{\mathrm{S}}\alpha_s\omega_s\right)^n\right] \\ &\qquad\times\cos\left(\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{s=1}^{\mathrm{S}}\alpha_s\phi_s\right) \\ &= \mathrm{V}\,.\sum_{\{\alpha_s\}}\left[\prod_{s=1}^{\mathrm{S}}\mathrm{J}_{\alpha_s}(m_s)\right].\left[\sum_{n=0}^{\mathrm{N}}c_n\,.\sum_{\{n_s\}}\left\{\frac{n\,!}{\prod_{s=0}^{\mathrm{S}}n_s\,!}\,.\,\omega_0^{n_0}.\prod_{s=1}^{\mathrm{S}}(\alpha_s\omega_s)^{n_s}\right\}\right] \\ &\qquad\times\cos\left(\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{s=1}^{\mathrm{S}}\alpha_s\phi_s\right). \qquad (19) \end{aligned}$$

where $\{n_s\}$ can take any set of (S+1) positive integral values or zero, such that $\sum_{s=0}^{\mathrm{S}}n_s=n$ and the summation $\sum_{\{n_s\}}$ is carried out over all such sets possible.

Expression (19) may be rewritten

$$\begin{aligned} &\sum_{n=0}^{\mathrm{N}}\left[c_n\sum_{\{n_s\}}\left\{n\,!\prod_{=0}^{\mathrm{S}}\left(\frac{\omega_s^{n_s}}{n_s\,!}\right).\sum_{\{a_s\}}\mathrm{V}\,.\prod_{s=1}^{\mathrm{S}}\{\alpha_s^{\nu_s}\mathrm{J}\alpha_s(m_e)\}\right.\right. \\ &\qquad\left.\left.\times\cos\left(\omega_0 t+\overline{\alpha-1}\,.\frac{\pi}{2}+\sum_{s=1}^{\mathrm{S}}\alpha_s\phi_s\right)\right\}\right]. \qquad (20) \end{aligned}$$

We have shown in Appendix 1 that

$$\alpha_s^{n_s}\mathrm{J}_{\alpha_s}(m_s)=\frac{1}{2^{n_s}}\sum_{\kappa_s=-n_s}^{n_s}u_{\kappa_s,\,n_s}\mathrm{J}_{\alpha_s+\kappa_s}(m_s),$$

$u_{\kappa_s,\,n_s}$ being a function of κ_s, n_s and m_s only.

Whence

$$\begin{aligned} \prod_{s=1}^{\mathrm{S}}\left\{\alpha_s^{n_s}\mathrm{J}_{\alpha_s}(m_s)\right\} &= \frac{1}{2^{n-n_0}}\prod_{s=1}^{\mathrm{S}}\left\{\sum_{\kappa_s=-n_s}^{n_s}u_{\kappa_s,\,n_s}\mathrm{J}_{\alpha_s+\kappa_s}(m_s)\right\} \\ &= \frac{1}{2^{n-n_0}}\sum_{\{\kappa_s\}}\left\{\prod_{s=1}^{\mathrm{S}}u_{\kappa_s,\,n_s}\mathrm{J}_{a_s+\kappa_s}(m_s)\right\}, \end{aligned}$$

the summation $\sum\limits_{\{\kappa_s\}}$ being carried out over all possible values of the set $\{\kappa_s\}$, bearing in mind that κ_s is integral and satisfies the condition

$$-n_s \leqslant \kappa_s \leqslant n_s.$$

Thus

$$\sum_{\{\alpha_s\}} \mathrm{V} \cdot \left[\prod_{s=1}^{\mathrm{S}} \left\{ \alpha_s^{n_s} \mathrm{J}_{\alpha_s}(m_s) \right\} . \cos\left(\omega_0 t + \overline{\alpha - 1} \,.\, \frac{\pi}{2} + \sum_{s=1}^{\mathrm{S}} \alpha_s \phi_s \right) \right]$$

$$= \sum_{\{\alpha_s\}} \mathrm{V} \cdot \frac{1}{2^{n-n_0}} \left[\sum_{\{\kappa_s\}} \left\{ \prod_{s=1}^{\mathrm{S}} u_{\kappa_s, n_s} \mathrm{J}_{\alpha_s+\kappa_s}(m_s) \right\} . \cos\left(\omega_0 t + \overline{\alpha - 1} \,.\, \frac{\pi}{2} \right.\right.$$

$$\left.\left. + \sum_{s=1}^{\mathrm{S}} \alpha_s \phi_s \right) \right]$$

$$= \frac{1}{2^{n-n_0}} \sum_{\{\kappa_s\}} \left[\prod_{s=1}^{\mathrm{S}} (u_{\kappa_s, n_s}) \,..\, \sum_{\{\alpha_s\}} \mathrm{V} \,.\, \prod_{s=1}^{\mathrm{S}} \{ \mathrm{J}_{\alpha_s+\kappa_s}(m_s) \} \right.$$

$$\left. \times \cos\left(\omega_0 t + \overline{\alpha - 1} \,.\, \frac{\pi}{2} + \sum_{s=1}^{\mathrm{S}} \alpha_s \phi_s \right) \right]$$

$$= \frac{1}{2^{n-n_0}} \sum_{\{\kappa_s\}} \left[\prod_{s=1}^{\mathrm{S}} (u_{\kappa_s, n_s}) \,.\, \sum_{\{\alpha_s\}} \mathrm{V} \,.\, \prod_{s=1}^{\mathrm{S}} \{ \mathrm{J}_{\alpha_s+\kappa_s}(m_s) \} \right.$$

$$\times \cos\left(\omega_0 t + \overline{\alpha + \kappa - 1} \,.\, \frac{\pi}{2} - \kappa \,.\, \frac{\pi}{2} - \sum_{s=1}^{\mathrm{S}} \kappa_s \phi_s + \sum_{s=1}^{\mathrm{S}} \overline{\alpha_s + \kappa_s} \,.\, \phi_s \right)$$

where

$$\kappa = \sum_{s=1}^{\mathrm{S}} \kappa_s.$$

This is equal to

$$\frac{1}{2^{n-n_0}} \sum_{\{\kappa_s\}} \left[\prod_{s=1}^{\mathrm{S}} (u_{\kappa_s, n_s}) \,.\, \sin\left(\omega_0 t - \kappa \frac{\pi}{2} - \sum_{s=1}^{\mathrm{S}} \kappa_s \phi_s + \sum_{s=1}^{\mathrm{S}} m_s \cos \phi_s \right) \right]. \qquad (21)$$

We see from Appendix 1 that

$$u_{\kappa_s, n_s} = (-1)^{n_s - \kappa_s} \,.\, u_{-\kappa_s, n_s},$$

whence
$$\prod_{s=1}^{\mathrm{S}} (u_{\kappa_s, n_s}) = (-1)^{n-n_0-\kappa} \,.\, \prod_{s=1}^{\mathrm{S}} (u_{-\kappa_s, n_s}).$$

Therefore, the expression (21) is

$$\frac{1}{2^{n-n_0+1}} \sum_{\{\kappa_s\}} \prod_{s=1}^{\mathrm{S}} (u_{\kappa_s, n_s}) \,.\, \left[\sin\left(\omega_0 t - \kappa \frac{\pi}{2} - \sum_{s=1}^{\mathrm{S}} \kappa_s \phi_s + \sum_{s=1}^{\mathrm{S}} m_s \cos \phi_s \right) \right.$$

$$\left. + (-1)^{n-n_0-\kappa} \sin\left(\omega_0 t + \kappa \frac{\pi}{2} + \sum_{s=1}^{\mathrm{S}} \kappa_s \phi_s + \sum_{s=1}^{\mathrm{S}} m_s \cos \phi_s \right) \right]$$

$$
\begin{aligned}
&= \frac{1}{2^{n-n_0+1}} \sum_{\{\kappa_s\}} \prod_{s=1}^{S} (u_{\kappa_s, n_s}) \cdot \left[\sin\left(\omega_0 t - \kappa\frac{\pi}{2} - \sum_{s=1}^{S} \kappa_s\phi_s + \sum_{s=1}^{S} m_s \cos\phi_s\right) \right. \\
&\qquad \left. + \sin\left(\omega_0 t + \kappa\frac{\pi}{2} + \sum_{s=1}^{S} \kappa_s\phi_s + \overline{n-n_0-\kappa}\,.\,\pi + \sum_{s=1}^{S} m_s \cos\phi_s\right)\right] \\
&= \frac{1}{2^{n-n_0}} \sum_{\{\kappa_s\}} \prod_{s=1}^{S} (u_{\kappa_s, n_s}) \cdot \cos\left(\sum_{s=1}^{S} \kappa_s\phi_s + \overline{n-n_0}\cdot\frac{\pi}{2}\right) \\
&\qquad \times \sin\left(\omega_0 t + \overline{n-n_0-\kappa}\,.\frac{\pi}{2} + \sum_{s=1}^{S} m_s \cos\phi_s\right) \\
&= \frac{1}{2^{n-n_0}} \sum_{\{\kappa_s\}}{}_1 \prod_{s=1}^{S} (u_{\kappa_s, n_s}) \cos\left(\sum_{s=1}^{S} \kappa_s\phi_s + \overline{n-n_0}\cdot\frac{\pi}{2}\right) . (-1)^{\frac{n-n_0-\kappa}{2}} \\
&\qquad \times \sin\left(\omega_0 t + \sum_{s=1}^{S} m_s \cos\phi_s\right) \\
&\quad + \frac{1}{2^{n-n_0}} \sum_{\{\kappa_s\}}{}_2 \prod_{s=1}^{S} (u_{\kappa_s, n_s}) \cos\left(\sum_{s=1}^{S} \kappa_s\phi_s + \overline{n-n_0}\cdot\frac{\pi}{2}\right) . (-1)^{\frac{n-n_0-\kappa-1}{2}} \\
&\qquad \times \cos\left(\omega_0 t + \sum_{s=1}^{S} m_s \cos\phi_s\right),
\end{aligned}
$$

where the summation $\sum_{\{\kappa_s\}}{}_1$ is carried out over all values of $\{\kappa_s\}$ for which $(n-n_0-\kappa)$ is even, and the summation $\sum_{\{\kappa_s\}}{}_2$ is carried out over all values of $\{\kappa_s\}$ for which $(n-n_0-\kappa)$ is odd.

Substituting this result in equation (21), we have

$$
\begin{aligned}
v_0 &= \mathrm{V}\,.\sum_{n=0}^{N} \left[c_n \,.\sum_{\{n_s\}} \frac{1}{2^{n-n_0}} \left\{ n\,! \prod_{s=0}^{S} \frac{\omega_s^{n_s}}{n_s\,!} \left[\sum_{\{\kappa_s\}}{}_1 \prod_{s=1}^{S} (u_{\kappa_s, n_s}) \right.\right.\right. \\
&\quad \times (-1)^{\frac{n-n_0-\kappa}{2}} \cos\left(\sum_{s=1}^{S} \kappa_s\phi_s + \overline{n-n_0}\cdot\frac{\pi}{2}\right) . \sin\left(\omega_0 t + \sum_{s=1}^{S} m_s \cos\phi_s\right) \\
&\quad + \sum_{\{\kappa_s\}}{}_2 \prod_{s=1}^{S} (u_{\kappa_s, n_s}) . (-1)^{\frac{n-n_0-\kappa-1}{2}} . \cos\left(\sum_{s=1}^{S} \kappa_s\phi_s + \overline{n-n_0}\cdot\frac{\pi}{2}\right) \\
&\quad \left.\left.\left. \times \cos\left(\omega_0 t + \sum_{s=1}^{S} m_s \cos\phi_s\right)\right]\right\}\right] \\
&= \Sigma_1' \cos\left(\omega_0 t + \sum_{s=1}^{S} m_s \cos\phi_s\right) + \Sigma_2' \sin\left(\omega_0 t + \sum_{s=1}^{S} m_s \cos\phi_s\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\Sigma_1'/\mathrm{V} &= \sum_{n=0}^{N} c_n \sum_{\{n_s\}} \frac{1}{2^{n-n_0}} \left\{ n\,! \prod_{s=0}^{S} \frac{\omega^{n_s}}{n_s\,!} \left[\sum_{\{\kappa_s\}}{}_2 \prod_{s=1}^{S} (u_{\kappa_s, n_s}) \right.\right. \\
&\qquad \left.\left. \times (-1)^{\frac{n-n_0-\kappa-1}{2}} . \cos\left(\sum_{s=1}^{S} \kappa_s\phi_s + \overline{n-n_0}\cdot\frac{\pi}{2}\right)\right]\right\}
\end{aligned}
$$

and

$$\Sigma_2'/\mathrm{V}=\sum_{n=0}^{\mathrm{N}} c_n \sum_{\{n_s\}} \frac{1}{2^{n-n_0}}\left\{n\,!\prod_{s=0}^{\mathrm{S}}\frac{\omega_s^{n_s}}{n_s!}\left[\sum_{\{\kappa_s\}}{}_1 \prod_{s=1}^{\mathrm{S}}(u_{\kappa_s,\,n_s})\right.\right.$$

$$\left.\left.\times(-1)^{\frac{n-n_0-\kappa}{2}}\,.\,\cos\left(\sum_{s=1}^{\mathrm{S}}\kappa_s\phi_s+\overline{n-n_0}\,.\frac{\pi}{2}\right)\right]\right\}.$$

We thus see that

$$v_0=\mathrm{V}\,.\,\sqrt{\Sigma_1'^2+\Sigma_2'^2}\,.\,\sin\left(\omega_0t+\sum_{s=1}^{\mathrm{S}}m_s\cos\phi_s+\tan^{-1}\frac{\Sigma_1'}{\Sigma_2'}\right). \quad . \quad (22)$$

The output E.M.F. is again shown to be a " hybrid " wave consisting of a carrier, modulated in frequency and amplitude by different distorted versions of the original applied signal.

It should be noted that the two functions of equations (7) and (8) representing the amplitude characteristic, will only represent the same curve over a finite range of angular frequencies W, and this representation can be made as exact as we please by taking a sufficient number of terms in each of the equations. Outside this range the equations will represent entirely different curves, that of equation (7) being periodic, and that of equation (8) being non-periodic.

It is of interest to consider the hypothetical case of a non-linear amplitude characteristic which has an exactly sinusoidal form, extending to infinite frequency (in both positive and negative directions). In this case the solution of Method I. is obviously exact (subject only to the limitations discussed in Section 6), only one Fourier term in equation (8) for the characteristic being required. However, the solution using Method II. will not be valid, since to represent a purely sinusoidal characteristic by a power series an infinite number of terms are required, and the appropriate series for the output components corresponding to equation (19) is not necessarily absolutely convergent, and hence cannot be subject to the analysis of Method II.

We should not, therefore, compare directly the results obtained by Methods I. and II. for the output from a network having some given amplitude-frequency characteristic, when a frequency modulated wave is applied.

6. " *Negative* " *Frequencies.*

We have discussed in Sections 2, 3, 4, and 5 the output from networks having specified amplitude-frequency and phase-shift-frequency characteristics. Thus, in Section 2, we take the amplitude-frequency characteristic to have the form of equation (2) and the phase-shift-frequency characteristic to be identically zero. No physical network could have such a characteristic because it would have to provide different attenuations for positive and negative frequencies of equal magnitude.

Strictly the linear network of Section 2 has a characteristic whose equation is

$$v_0=v_i(c_0+c_1\,|\,\omega\,|\,). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (23)$$

The output from the network of equation (23) is substantially equal to that from equation (2) if the sidebands in equation (4), for which $\left(\omega_0+\sum\limits_{s=1}^{S}\alpha_s\omega_s\right)$ is negative, have amplitudes which are very small.

Thus, the output from a network having a linear amplitude characteristic is a substantially undistorted "hybrid" wave, only if the carrier frequency is large compared with the modulating frequencies.

However, the form given in equation (6) for the linear phase-shift-frequency characteristic is physically realizable, and the output from a network having such a characteristic is an undistorted "hybrid" wave even when the carrier frequency is not large compared with the modulating frequencies.

Applying similar considerations to the discussion of Section 5, we see that the analysis there given is again strictly true only if the carrier frequency is large compared with the modulating frequencies. In each case "large" means so large that the calculated amplitudes of the components of negative frequency in the output are small. However, it is worth noting that in Section 5 (b), if the amplitude-frequency characteristic is represented by a series of cosine terms, thus:

$$v_0=v_i\sum_{n=1}^{N}g_n\cos\frac{2\pi n\omega}{W}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (24)$$

and if in Section 5 (d) it is represented by a polynomial consisting entirely of even powers, thus:

$$v_0=v_i\sum_{n=0}^{N}c_n\omega^{2n}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (25)$$

then this condition of carrier frequency large compared with modulating frequencies is no longer necessary. Indeed, in the consideration of any problem in which the assumption of carrier frequency large is not allowable, the analytical expression which is fitted to the experimental amplitude-frequency characteristic should take one or other of the forms (24) and (25).

Again, if in Section 5 (c), the phase-shift-frequency characteristic is represented by a series of sine terms, thus:

$$\psi_0=\eta_0+\eta_1\omega+\sum_{n=1}^{N}\nu_n\sin\frac{2\pi n\omega}{W}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (26)$$

then we do not need to make the assumption of carrier frequency large. If in any problem the carrier frequency is not large, then an analytical expression of the form (26) should be fitted to the characteristic.

APPENDIX 1.

Expansion of $(2\alpha)^r \,.\, J_\alpha(m)$.

It may easily be seen that by repeated application of the formula

$$2\alpha \,.\, J_\alpha(m) = m[J_{\alpha+1}(m) + J_{\alpha-1}(m)], \quad \ldots \ldots \quad (27)$$

$(2\alpha)^r J_\alpha(m)$ may be expanded as the series

$$\sum_{i=-r}^{r} u_{i,r} \,.\, J_{\alpha+i}(m),$$

where $u_{i,r}$ are coefficients independent of α. For suppose

$$(2\alpha)^r \,.\, J_\alpha(m) = \sum_{i=-r}^{r} u_{i,r} J_{\alpha+i}(m),$$

then

$$\begin{aligned} (2\alpha)^{r+1} \,.\, J_\alpha(m) &= \sum_{i=-r}^{r} u_{i,r} \,.\, 2\alpha J_{\alpha+i}(m) \\ &= \sum_{i=-r}^{r} u_{i,r}[2(\alpha+i)J_{\alpha+i}(m) - 2iJ_{\alpha+i}(m)] \\ &= \sum_{i=-r}^{r} u_{i,r}[m\{J_{\alpha+i-1}(m) + J_{\alpha+i+1}(m)\} - 2iJ_{\alpha+i}(m)]. \quad (28) \end{aligned}$$

But

$$(2\alpha)^{r+1} \,.\, J_\alpha(m) = \sum_{i=-(r+1)}^{r+1} u_{i,r+1} J_{\alpha+i}(m). \quad \ldots \quad (29)$$

Combining the equations (28) and (29), we have the identity

$$\sum_{i=-(r+1)}^{r+1} u_{i,r+1} J_{\alpha+1}(m) \equiv \sum_{i=-r}^{r} u_{i,r}[m\{J_{\alpha+i-1}(m) + J_{\alpha+i+1}(m)\} - 2iJ_{\alpha+i}(m)].$$

Whence, equating coefficients of the Bessel functions of equal argument, we have

$$u_{\pm(r+1),(r+1)} = mu_{\pm r,r}, \quad \ldots \ldots \quad (30)$$

$$u_{\pm r,(r+1)} = mu_{\pm(r-1),r} \mp 2ru_{r,r}, \quad \ldots \ldots \quad (31)$$

and

$$u_{\pm i,(r+1)} = m[u_{\pm(i+1),r} + u_{\pm(i-1),r}] \mp 2iu_{\pm i,r}. \quad \ldots \quad (32)$$

for all integral values of i from 0 to $(r-1)$ inclusive.

From equation (22) we see that

$$u_{\pm 1,1} = m \quad \text{and} \quad u_{0,1} = 0.$$

By repeated application of the equations (30), (31), and (32), we can work out the values of the coefficients $u_{i,r}$ for any desired values of r and i $(\leqslant r)$.

The values of these coefficients, for all values of r from 1 to 8 inclusive and all allowable values of i, are given in the Table.

APPENDIX 2.

The Convergence of the Series $\sum\limits_{\alpha=0}^{\infty} \alpha^n J_\alpha(m) \cos \alpha\phi$.

In the series $\sum\limits_{\alpha=0}^{\infty} |\alpha^n J_\alpha(m)|$, the ratio of the αth term to the $(\alpha+1)$th term is $\left(\frac{\alpha+1}{\alpha}\right)^n . \frac{J_{\alpha+1}(m)}{J_\alpha(m)}$. Now, as $\alpha \rightarrow \infty$, $J_\alpha(m) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\alpha \log \alpha}$.

Hence
$$\left|\left(\frac{\alpha+1}{\alpha}\right)^n . \frac{J_{\alpha+1}(m)}{J_\alpha(m)}\right| \rightarrow \left|\left(\frac{\alpha+1}{\alpha}\right)^n \frac{e^{\alpha \log \alpha}}{e^{(\alpha+1)\log(\alpha+1)}}\right|,$$
which tends to zero as $\alpha \rightarrow \infty$.

Thus the series $\sum\limits_{\alpha=0}^{\infty} |\alpha^n J_\alpha(m)|$ is convergent, and so $\sum\limits_{\alpha=0}^{\infty} \alpha^n J_\alpha(m)$ is absolutely convergent.

Now $|\cos \alpha\phi| \leqslant 1$. Hence $\sum\limits_{\alpha=0}^{\infty} \alpha^n . J_\alpha(m) . \cos \alpha\phi$ is absolutely convergent.

APPENDIX 3.

The Convergence of the Series $\sum\limits_{\{\alpha_s\}} \left[\prod\limits_{s=1}^{S} \alpha_s^{n_s} J_{\alpha_s}(m_s)\right] \times \cos\left(\omega_0 t + \overline{\alpha-1} . \frac{\pi}{2} + \sum\limits_{s=1}^{S} \alpha_s \phi_s\right)$.

Consider the product
$$\prod_{s=1}^{S} \left[\sum_{\alpha_s=-\infty}^{\infty} \alpha_s^{n_s} J_{\alpha_s}(m_s)\right], \qquad (33)$$
where n_s is a positive integer.

From Appendix 2, it may be seen that each of the series
$$\sum_{\{\alpha_s\}=-\infty}^{\infty} \alpha_s^{n_s} . J_{\alpha_s}(m_s)$$
is absolutely convergent, so that the expression (33) may be expanded in the form of the absolutely convergent series
$$\sum_{\{\alpha_s\}} \left[\prod_{s=1}^{S} \alpha_s^{n_s} J_{\alpha_s}(m_s)\right].$$

Since
$$\left|\cos\left(\omega_0 t + \overline{\alpha-1} . \frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s \phi_s\right)\right| \leqslant 1;$$

r i	1.	2.	3.	4.	5.	6.	7.	8.
8								m^8
7							m^7	$-56m^7$
6						m^6	$-42m^6$	$8m^8+1064m^6$
5					m^5	$-30m^5$	$7m^7+560m^5$	$-280m^7-8400m^5$
4				m^4	$-20m^4$	$6m^6+260m^4$	$-168m^6-2800m^4$	$28m^8+3024m^6+27216m^4$
3			m^3	$-12m^3$	$5m^5+100m^3$	$-90m^5-720m^3$	$21m^7+1120m^5+4816m^3$	$-504m^7-11760m^5-30912m^3$
2		m^2	$-6m^2$	$4m^4+28m^2$	$-40m^4-120m^2$	$15m^6+320m^4+496m^2$	$-210m^6-2240m^4=2016m^2$	$56m^8+2520m^6+14784m^4+8128m^2$
1	m	$-2m$	$3m^3+4m$	$-12m^3-8m$	$10m^5+60m^3+16m$	$-60m^5-240m^3-32m$	$35m^7+560m^5+1008m^3+64m$	$-280m^7-3360m^5=4032m^3-128m$
0	o	$2m^2$	0	$6m^4+8m^2$	0	$20m^6+120m^4+32m^2$	0	$70m^8+1120m^6+2016m^4+128m^2$
-1	m	$2m$	$3m^3+4m$	$12m^3+8m$	$10m^5+60m^3+16m$	$60m^5+240m^3+32m$	$35m^7+560m^5+1008m^3+64m$	$280m^7+3360m^5+4032m^3+128m$
-2		m^2	$6m^2$	$4m^4+28m^2$	$40m^4+120m^2$	$15m^6+320m^4+496m^2$	$210m^6+2240m^4+2016m^2$	$56m^8+2520m^6+14784m^4+8128m^2$
-3			m^3	$12m^3$	$5m^5+100m^3$	$90m^5+720m^3$	$21m^7+1120m^5+4816m^3$	$504m^7+11760m^5+30912m^3$
-4				m^4	$20m^4$	$6m^6+260m^4$	$168m^6+2800m^4$	$28m^8+3024m^6+27216m^4$
-5					m^5	$30m^5$	$7m^7+560m^5$	$280m^7+8400m^5$
-6						m^6	$42m^6$	$8m^8+1064m^6$
-7							m^7	$56m^7$
-8								m^8

the series

$$\sum_{\{\alpha_s\}} \left[\prod_{s=1}^{S} \alpha_s^{n_s} J_{\alpha_s}(m_s) \right] . \cos\left(\omega_0 t + \overline{\alpha - 1} . \frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s \phi_s \right)$$

is absolutely convergent.

References.

(1) E. C. Cherry and R. S. Rivlin, "Non-Linear Distortion, with Particular reference to the Theory of Frequency Modulated Waves.—Part I." Phil. Mag. vol. xxxii. p. 265 (1941).
(2) H. Roder, "Effect of Tuned Circuits on a Frequency Modulated Signal." Proc. I. R. E. vol. xxv. p. 1617 (1937).
(3) J. Carson and T. Fry, "Variable Frequency Electric Circuit Theory." Bell Syst. Tech. J. vol. xvi. p. 513 (1937).

XXIX. *Note on the Bending of Thin Uniformly Loaded Plates bounded by Cardioids, Lemniscates, and certain other Quartic Curves.*

By Bibhutibhusan Sen, Islamia College, Bengal *.

[Received February 9, 1940.]

Summary.

In this note the bending of thin uniformly loaded plates bounded by cardioids, lemniscates and certain other quartic curves is discussed. Assuming that the boundary is clamped, the solution in each case is obtained in a finite number of terms.

Introduction.

The normal displacement w of a thin plate bent by transverse forces, satisfies the differential equation †

$$\frac{\partial^4 w}{\partial x^4}+2\frac{\partial^4 w}{\partial x^2\,\partial y^2}+\frac{\partial^4 w}{\partial y^4}=\frac{Z}{D}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (A)$$

where Z=the constant load per unit of area,
and D=the flexural rigidity.

For a clamped edge, we have on the boundary

$$w=\frac{\partial w}{\partial \nu}=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (B)$$

$\partial \nu$ being an element of the outward drawn normal.

The exact solution of the equation (A) satisfying the boundary conditions (B) is known only for some thin plates of simple forms. The object of this paper is to consider the same problem when the boundaries are cardioids, lemniscates, inverses of ellipses and elliptic limacons. The results are interesting inasmuch as they are expressed in a finite number of terms.

1. *Solution for a Plate bounded by a Cardioid.*

By using the transformation

$$\xi+i\eta=\frac{1}{(x-iy)^{\frac{1}{2}}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.1)$$

---

* Communicated by the Author.

† ‘The Mathematical Theory of Elasticity,’ by A. E. H. Love, 3rd edition, p. 492 (1920).

we obtain

$$\left.\begin{aligned} &\xi = r^{-\frac{1}{2}} \cos\frac{\theta}{2}, \quad \eta = r^{-\frac{1}{2}} \sin\frac{\theta}{2}, \\ \text{and} \quad &\frac{1}{h^2} = \left(\frac{\partial x}{\partial \xi}\right)^2 + \left(\frac{\partial y}{\partial \xi}\right)^2 = \frac{4}{(\xi^2+\eta^2)^3} \end{aligned}\right\} \quad . \quad . \quad . \quad . \qquad \mathbf{(1.2)}$$

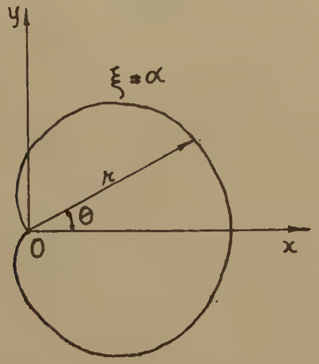
It is evident that ξ=constant and η=constant give two families of orthogonal cardioids. We shall take as the boundary of the plate

$$\xi = \alpha = (2a)^{-\frac{1}{2}}, \quad \text{or} \quad r = a(1+\cos\theta). \quad . \quad . \quad . \quad . \qquad (1.3)$$

The above transformation enables us to write the equation (A) in the form

$$\left[(\xi^2+\eta^2)^3\left(\frac{\partial^2}{\partial \xi^2}+\frac{\partial^2}{\partial \eta^2}\right)\right]^2 w = \frac{16\mathrm{Z}}{\mathrm{D}} = 16 \times 64\, p_0, \quad . \quad . \quad . \qquad (1.4)$$

Fig. 1.



where

$$p_0 = \frac{\mathrm{Z}}{64\mathrm{D}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad \mathbf{(1.5)}$$

The particular integral of this equation is given by

$$w = \frac{p_0}{(\xi^2+\eta^2)^4}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad \mathbf{(1.6)}$$

Combining this expression with some suitable terms which make

$$\left[(\xi^2+\eta^2)^3\left(\frac{\partial^2}{\partial \xi^2}+\frac{\partial^2}{\partial \eta^2}\right)\right]^2 w = 0, \quad . \quad . \quad . \quad . \quad . \qquad (1.7)$$

we obtain

$$w = \frac{p_0 + \mathrm{A}(\xi^2-\eta^2) + \mathrm{B}\xi^2\eta^2}{(\xi^2+\eta^2)^4} + \frac{\mathrm{C}\xi + \mathrm{D}\xi^3}{(\xi^2+\eta^2)^3}, \quad . \quad . \quad . \quad . \qquad (1.8)$$

where A, B, C, D are constants.

From the boundary conditions (B) we get

$$w=\frac{\partial w}{\partial \xi'}=0, \quad \text{when} \quad \xi=\alpha.$$

These relations give

$$\left.\begin{aligned} A&=-\frac{p_0}{\alpha^2}, \\ B&=-\frac{p_0}{\alpha^4}, \\ C&=-\frac{p_0}{\alpha^3}, \\ D&=\frac{p_0}{\alpha^5}. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.9)$$

When these values of the constants are substituted in (1.8), we have

$$w=\frac{p_0}{(\xi^2+\eta^2)^4}\left[1-\frac{\xi^2}{\alpha^2}-\frac{\xi^3}{\alpha^3}+\frac{\xi^5}{\alpha^5}+\frac{\eta^2}{\alpha^2}\left(1-\frac{\xi}{\alpha}-\frac{\xi^2}{\alpha^2}+\frac{\xi^3}{\alpha^3}\right)\right]. \quad (1.10)$$

Expressed in polar co-ordinates, this result stands as

$$\begin{aligned} w=p_0r^4\Bigg[1&-\left(\frac{2a}{r}\right)\cos^2\frac{\theta}{2}-\left(\frac{2a}{r}\right)^{3/2}\cos^3\frac{\theta}{2}+\left(\frac{2a}{r}\right)^{5/2}\cos^5\frac{\theta}{2} \\ &+\left(\frac{2a}{r}\right)\sin^2\frac{\theta}{2}\left\{1-\left(\frac{2a}{r}\right)^{1/2}\cos\frac{\theta}{2}-\left(\frac{2a}{r}\right)\cos^2\frac{\theta}{2}+\left(\frac{2a}{r}\right)^{3/2}\cos^3\frac{\theta}{2}\right\}\Bigg]. \quad (1.11) \end{aligned}$$

It is apparent that the above expression for the normal displacement is symmetrical about the line $\theta=0$. The value of the deflexion along this line is given by

$$(w)_{\theta=0}=p_0\rho^4\left[1-\frac{2a}{\rho}-\left(\frac{2a}{\rho}\right)^{3/2}+\left(\frac{2a}{\rho}\right)^{5/2}\right], \quad . \quad . \quad (1.12)$$

where ρ is the distance measured along the line of symmetry from the origin.

On putting

$$\left(\frac{\rho}{2a}\right)^{1/2}=X, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.13)$$

the above expression becomes

$$(w)_{\theta=0}=p_0(2a)^4X^3(X^3-1)(X^2-1). \quad . \quad . \quad . \quad . \quad (1.14)$$

By simple calculation we find that the maximum displacement on the line of symmetry occurs at a distance

$$\rho=\cdot 908a \text{ (nearly)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.15)$$

where its approximate value is

$$\frac{Z}{D}(\cdot 029)a^4. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (1.16)$$

2. *Solution for a Plate bounded by a Lemniscate of Bernoulli.*

Let us consider the system of orthogonal co-ordinates (ξ, η) defined by

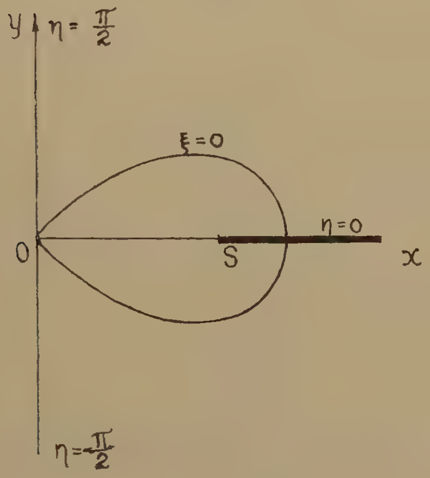
$$\left.\begin{array}{l} (z+c)(z-c)=c^2e^{2\zeta}, \\ \zeta=\xi+i\eta \quad \text{and} \quad z=x+iy, \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \qquad (2.1)$$

where

In cartesian co-ordinates we obtain the relations

$$\left.\begin{array}{l} (x^2-y^2-c^2)^2+4x^2y^2=c^4e^{4\xi}, \\ (x^2-y^2-c^2)=2xy \cot 2\eta \end{array}\right\}, \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2.2)$$

Fig. 2.



so that ξ=constant gives a family of quartic curves * with real foci at $z=\pm c$ and η=constant gives the orthogonal family of curves which are rectangular hyperbolas, each of which passes through the foci of the quartic curves. From (2.2) it is evident that the curve $\xi=0$ gives a lemniscate of Bernoulli of which the polar equation is

$$r^2=2c^2 \cos 2\theta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2.3)$$

At infinity $\xi \to \infty$ while at either focus $\xi \to -\infty$. We shall consider the loop of the lemniscate of Bernoulli given by the curve $\xi=0$, which lies in the region

---

* Basset, 'Treatise on Hydrodynamics,' vol. i. p. 106 (1888).

$$-\infty \leqslant \xi \leqslant 0, \quad -\frac{\pi}{2} \leqslant \eta \leqslant \frac{\pi}{2}.$$

Since, in the present case,

$$\frac{1}{h^2}=\left(\frac{\partial x}{\partial \xi}\right)^2+\left(\frac{\partial y}{\partial \xi}\right)^2=\frac{c^2e^{3\xi}}{\sqrt{2}(\cosh 2\xi+\cos 2\eta)^{\frac{1}{2}}}, \quad . \quad . \quad . \quad . \quad (2.4)$$

the equation (A) can be expressed in terms of ξ and η in the form

$$\left[\frac{(\cosh 2\xi+\cos 2\eta)^{\frac{1}{2}}}{e^{3\xi}}\left(\frac{\partial^2}{\partial \xi^2}+\frac{\partial^2}{\partial \eta^2}\right)\right]^2 w=\frac{c^4\mathrm{Z}}{2\mathrm{D}}=32p_0, \quad . \quad . \quad (2.5)$$

where

$$p_0=\frac{c^4\mathrm{Z}}{64\mathrm{D}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.6)$$

The particular integral of the above equation will be

$$w=p_0e^{4\xi}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.7)$$

Finding out a few terms which make the left-hand side of (2.5) equal to zero, we can write for the purpose of our problem

$$w=p_0e^{4\xi}+\mathrm{P}_1+\mathrm{P}_2e^{2\xi}\cos 2\eta+\mathrm{P}_3e^{2\xi}\cos \eta\,(\cosh 2\xi+\cos 2\eta)^{\frac{1}{2}}, \qquad (2.8)$$

where P_1, P_2, and P_3 are constants.

For a clamped edge we must have on the boundary

$$w=0, \quad \frac{\partial w}{\partial \xi}=0, \quad \text{when} \quad \xi=0. \quad . \quad . \quad . \quad . \quad . \quad (2.9)$$

These conditions are satisfied if

$$\begin{aligned}\mathrm{P}_1&=p_0,\\ \mathrm{P}_2&=2p_0, \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.10)\\ \mathrm{P}_3&=-2\sqrt{2}p_0.\end{aligned}$$

Hence the normal displacement in this case is obtained as

$$w=p_0[e^{4\xi}+1+2e^{2\xi}\{\cos 2\eta-\sqrt{2}\cos \eta\,(\cosh 2\xi+\cos 2\eta)^{\frac{1}{2}}\}]. \quad (2.11)$$

At the focus where $\xi \rightarrow -\infty$,

$$w=p_0=\frac{c^4\mathrm{Z}}{64\mathrm{D}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.12)$$

3. *Solution for a Plate bounded by an Inverse of an Ellipse.*

If we use the transformation *

$$x+iy=c\sec(\xi+i\eta) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.1)$$

---

* Basset, *loc. cit.* p. 102.

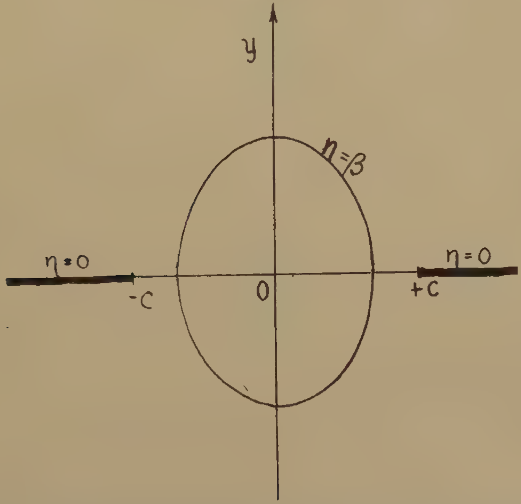
we get

$$\left.\begin{aligned} x&=\frac{2c\cosh\eta\cos\xi}{\cosh 2\eta+\cos 2\xi}, \qquad y=\frac{2\,c\sinh\eta\sin\xi}{\cosh 2\eta+\cos 2\xi}, \\ r^4&=c^2\left(\frac{x^2}{\cosh^2\eta}+\frac{y^2}{\sinh^2\eta}\right)=c^2\left(\frac{x^2}{\cos^2\xi}-\frac{y^2}{\sin^2\xi}\right), \end{aligned}\right\} \quad . \quad . \quad (3.2)$$

and
$$\frac{1}{h^2}=\left(\frac{\partial x}{\partial\xi}\right)^2+\left(\frac{\partial y}{\partial\xi}\right)^2=\frac{2c^2(\cosh 2\eta-\cos 2\xi)}{(\cosh 2\eta+\cos 2\xi)^2}. \quad . \quad . \quad . \quad (3.3)$$

It is evident that $\eta=\beta$ is the inverse of an ellipse and $\xi=\alpha$ is an inverse of a confocal hyperbola which becomes an arc of a lemniscate of Bernoulli when $\xi=\frac{\pi}{4}$. Here ξ may have any real value positive or negative, but

Fig. 3.



as the values of x and y are periodic in ξ, it is only necessary to consider values of ξ lying between 0 and 2π. Values of η are restricted to all positive real values between zero and infinity. At the origin $\eta\rightarrow\infty$ and when $\eta=0$, we have two double lines extending from the points $x=\pm c$ to infinity in the positive and negative directions respectively.

In the present case the equation (A) reduces to

$$\left[\frac{(\cosh 2\eta+\cos 2\xi)^2}{\cosh 2\eta-\cos 2\xi}\left(\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\eta^2}\right)\right]^2 w=\frac{4c^4\mathrm{Z}}{\mathrm{D}}=64p \text{ (say)}. \quad . \quad (3.4)$$

For clamped edge the boundary conditions are

$$\left.\begin{aligned} w&=0, \\ \frac{\partial w}{\partial\eta}&=0 \end{aligned}\right\} \text{ when } \eta=\beta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.5)$$

for an inverse of an ellipse. For an inverse of a hyperbola we shall have

$$\left.\begin{aligned} w&=0,\\ \frac{\partial w}{\partial \xi}&=0 \end{aligned}\right\} \text{ when } \xi=\alpha. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (3.6)$$

As the solution of the equation (3.4) we can take *

$$w=\frac{p+\mathrm{P}_1(\cosh 2\eta\cos 2\xi+1)+\mathrm{P}_2\sinh 2\eta}{(\cosh 2\eta+\cos 2\xi)^2}+\frac{\mathrm{P}_3\sinh 2\eta+\mathrm{P}_4}{(\cosh 2\eta+\cos 2\xi},\qquad (3.7)$$

where P_1, P_2, P_3, P_4 are constants.

The boundary conditions (3.5) will be satisfied if

$$\begin{aligned} \mathrm{P}_1&=-p\operatorname{cosech}^2 2\beta,\\ \mathrm{P}_2&=-2p\operatorname{cosech} 2\beta,\\ \mathrm{P}_3&=\ \ 2p\operatorname{cosech} 4\beta,\\ \mathrm{P}_4&=\ \ 2p\operatorname{cosech} 2\beta\operatorname{cosech} 4\beta \end{aligned} \quad . \quad . \quad . \quad . \quad . \quad . \qquad (3.8)$$

At the centre where $\eta\to\infty$, we have

$$\left.\begin{aligned} w&=2p\operatorname{cosech} 4\beta,\\ &=\frac{c^4\mathrm{Z}}{8\mathrm{D}}\operatorname{cosech} 4\beta \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (3.9)$$

4. *Solution for a Plate bounded by an Elliptic Limacon.*

By using the transformation †

$$\xi+i\eta=2\sec^{-1}\sqrt{\frac{x+iy}{2c}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (4.1)$$

we find that $\eta=\beta$ (constant) represents an inverse of an ellipse with respect to the focus as the pole, that is, an elliptic limacon. The values of η

---

* For the inverse of a hyperbola we can take the solution of (3.4) as

$$w=\frac{p+\mathrm{Q}_1(\cosh 2\eta\cos 2\xi+1)+\mathrm{Q}_2\sin 2\xi}{(\cosh 2\eta+\cos 2\xi)^2}+\frac{\mathrm{Q}_3\sin 2\xi+\mathrm{Q}_4}{(\cosh 2\eta+\cos 2\xi)},\qquad (3.7')$$

where Q_1, Q_2, Q_3, Q_4 are constants.

Provided α be not equal to $\pi/4$ or a multiple of $\pi/4$, the boundary condition (3.6) will give

$$\left.\begin{aligned} \mathrm{Q}_1&=p\operatorname{cosec} 2\alpha^2, & \mathrm{Q}_2&=-2p\operatorname{cosec} 2\alpha,\\ \mathrm{Q}_3&=2p\operatorname{cosec} 4\alpha, & \mathrm{Q}_4&=-2p\operatorname{cosec} 2\alpha\operatorname{cosec} 4\alpha \end{aligned}\right\} \quad . \quad . \qquad (3.8')$$

Special solutions can also be obtained when $\xi=\pi/4$ or a multiple of $\pi/4$.

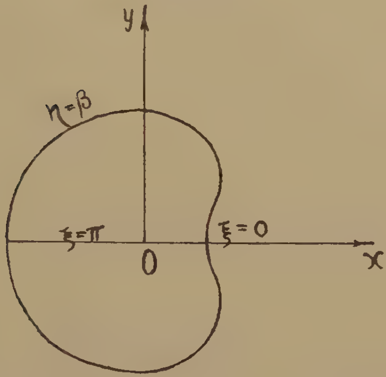
† Basset, *loc. cit.* p. 106.

are restricted to all real positive values between zero and infinity, while ξ may have any real value positive or negative.

From the above transformation we get

$$\left.\begin{aligned} \frac{4cx}{r^2}&=1+\cosh\eta\cos\xi,\\ \frac{4cy}{r^2}&=\sinh\eta\sin\xi,\\ r&=\frac{4c}{\cosh\eta+\cos\xi},\\ \text{and}\qquad \frac{1}{h^2}&=\left(\frac{\partial x}{\partial\xi}\right)^2+\left(\frac{\partial y}{\partial\xi}\right)^2=\frac{16c^2(\cosh\eta-\cos\xi)}{(\cosh\eta+\cos\xi)^3} \end{aligned}\right\} \quad . \quad . \quad . \quad (4.2)$$

Fig. 4.



These relations enable us to write the equation (A) in the form

$$\left[\frac{(\cosh\eta+\cos\xi)^3}{\cosh\eta-\cos\xi}\left(\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\eta^2}\right)\right]^2 w=\frac{256c^4Z}{D}=64\,p\ (\text{say}). \quad . \quad (4.3)$$

The solution of this equation suitable for the present problem can be obtained as

$$w=\frac{p+P_1(\cosh^2\eta\cos^2\xi-1)+P_2(\cosh\eta\cos\xi+1)}{(\cosh\eta+\cos\xi)^4}+\frac{P_3\sinh\eta+P_4\sinh 3\eta}{(\cosh\eta+\cos\xi)^3}+\frac{P_5\sinh 2\eta+P_6}{(\cosh\eta+\cos\xi)^2}, \quad (4.4)$$

P_1, P_2, P_3, P_4, P_5, P_6 being constants.

On the boundary $\eta=\beta$, we have

$$w=0, \quad \frac{\partial w}{\partial\eta}=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.5)$$

These conditions are satisfied if

$$\left.\begin{aligned}
P_1&=p \operatorname{cosech}^4 \beta,\\
P_2&=2P_1 \cosh^2 \beta,\\
P_3&=-\frac{P_1}{2}\coth \beta(4 \sinh^2 \beta+3),\\
P_4&=\frac{P_1}{2}\coth \beta,\\
P_5&=-\frac{P_1}{2}\tanh 2\beta,\\
P_6&=-P_1 \cosh^2 \beta \operatorname{sech} 2\beta
\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (4.6)$$

At the origin where $\eta\rightarrow\infty$, we have

$$w=\frac{4c^4Z}{D}\operatorname{cosech}^4 \beta[2 \coth \beta-\tanh 2\beta]. \quad . \quad . \quad . \quad (4.7)$$

XXX. *Heat Conduction in a Semi-infinite Slab.*

By W. A. Mersman [*].

[Received July 10, 1941.]

1. *Introduction.*

The problem to be solved here is that of heat conduction in a semi-infinite slab whose initial temperature distribution is known, when the temperature gradient at the face is a known function of the time. In § 2 a purely formal solution is obtained, using Laplace transformations and Riemann integrals. Then, substituting Lebesgue integrals, the solution is completely established in § 3. In § 4 is shown the connexion with the classical theory. Finally, in §§ 5, 6, two special cases are treated in detail, and methods of numerical application are indicated.

2. *The Formal Solution.*

We have the boundary-value problem

$$\frac{\partial u}{\partial t}=a\frac{\partial^2 u}{\partial x^2}, \quad (t>0,\ x>0), \quad \dots \dots \quad (1)$$

$$u(x,t)\to f(x) \text{ as } t\to+0, \quad (x>0), \quad \dots \dots \quad (2)$$

$$\frac{\partial u}{\partial x}\to g(t) \text{ as } x\to+0, \quad (t>0), \quad \dots \dots \quad (3)$$

where $u(x, t)$ is the temperature at point x at time t, a is the diffusivity, and $f(x)$ and $g(t)$ are known functions.

Let U, G be the Laplace transforms [†] of u, g respectively. Then equations (1)–(3) yield the transformed equations

$$\mathrm{SU}-f(x)=a\frac{\partial^2 \mathrm{U}}{\partial x^2} \quad (x>0), \quad \dots \dots \quad (4)$$

$$\frac{\partial u}{\partial x}\to \mathrm{G}(s) \text{ as } x\to+0. \quad \dots \dots \quad (5)$$

The general solution of (4), subject to the condition that U be bounded as $x\to\infty$, can be written in the form

$$\mathrm{U}(x,s)=\mathrm{A}\exp\left[-x(s/a)^{1/2}\right]+1/2(as)^{1/2}\int f(\xi)\exp\left[-\,|x-\xi|\,(s/a)^{1/2}\right]d\xi, \quad (6)$$

---

[*] Communicated by the Author.

[†] *Cf.* G. Doetsch, 'Theorie und Anwendung der Laplace Transformation,' Chap. 3.

where A is an arbitrary constant. Solving the boundary condition (5) for A, we obtain the desired particular solution

$$U(x, s)=-(a/s)^{1/2}G(s)\exp[-x(s/a)^{1/2}]$$
$$+1/2(as)^{1/2}\int_0^\infty f(\xi)\{\exp[-|x+\xi|(s/a)^{1/2}]+\exp[-|x-\xi|(s/a)^{1/2}]\}d\xi. \quad . \quad . \quad . \quad (7)$$

Applying the inverse Laplace transformations * we obtain, as a formal solution of the system (1)–(3)

$$u(x, t)=1/2(\pi at)^{1/2}\int_0^\infty f(\xi)\{\exp[-(x+\xi)^2/4at]+\exp[-(x-\xi)^2/4at]\}d\xi$$
$$-(a/\pi)^{1/2}\int_0^t g(\tau)(t-\tau)^{-1/2}\exp[-x^2/4a(t-\tau)]d\tau, \quad (x>0, t>0). \quad (8)$$

3. *The Solution Established.*

Write v for the first term of the right member of (8) and w for the second term. Then v is known † to be the solution of the differential equation (1), subject to the conditions

$$v(x, t)\to f(x) \text{ as } t\to+0, \quad (x>0), \quad . \quad . \quad . \quad . \quad . \quad (9)$$

$$\frac{\partial v}{\partial x}\to 0 \text{ as } x\to+0, \quad (t>0). \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Hence, we need to prove that the differential equation (1) and the conditions

$$w(x, t)\to 0 \text{ as } t\to+0, \quad (x>0), \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

$$\frac{\partial w}{\partial x}\to g(t) \text{ as } x\to+0, \quad (t>0), \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

are satisfied by the function

$$w(x, t)=-(a/\pi)^{1/2}\int_0^t g(\tau)(t-\tau)^{-1/2}\exp[-x^2/4a(t-\tau)]d\tau. \quad . \quad (13)$$

We shall prove this, subject to the restrictions that $g(\tau)$ be a bounded Lebesgue measurable function in the closed interval $[0, t]$ and continuous on the left at t, and the integral in (13), as well as all integrals following, is understood to be a Lebesgue integral.

---

* *Cf.* Doetsch, *ibid.*, Chap. 8 and the table at the end of the book.

† *Cf.* H. S. Carslaw, 'Introduction to the Mathematical Theory of the Conduction of Heat in Solids,' 2nd edition, Chap. 3, §§ 16–18, where sufficient conditions on $f(x)$ are given.

We begin with equation (12). Differentiating (13) with respect to x inside the integral sign gives

$$\frac{\partial w}{\partial x}=x/2(a\pi)^{1/2}\int_0^t g(\tau)(t-\tau)^{-3/2}\exp\left[-x^2/4a(t-\tau)\right]d\tau.$$

This is permissible, since the coefficient of $g(\tau)$ in the integrand is a continuous function of x uniformly in τ in $(0\leqq\tau\leqq t)$, as long as $x>0$ *. Now change the variable of integration by means of the substitution †

$$\xi^2=x^2/4a(t-\tau),$$

obtaining

$$\frac{\partial w}{\partial x}=2\pi^{-1/2}\int_y^\infty e^{-\xi^2}g(t-x^2/4a\xi^2)\,d\xi,\quad y=x/2(at)^{1/2}. \quad . \quad . \quad (14)$$

Since the argument of g is still restricted to the range $[0, t]$, this infinite Lebesgue integral is convergent. Observing that

$$2\pi^{-1/2}\int_0^\infty g(t)e^{-\xi^2}d\xi=g(t),$$

we can write

$$\frac{\partial w}{\partial x}-g(t)=2\pi^{-1/2}\int_y^{y+\delta}g(t-x^2/4a\xi^2)e^{-\xi^2}d\xi-2\pi^{-1/2}\int_0^{y+\delta}g(t)e^{-\xi^2}d\xi$$
$$+2\pi^{-1/2}\int_{y+\delta}^\infty[g(t-x^2/4a\xi^2)-g(t)]e^{-\xi^2}d\xi. \quad (15)$$

Since $g(\xi)$ is bounded in $[0, t]$, we can choose a positive number M such that $|g(\tau)|<\mathrm{M}$, $0\leqq\tau\leqq t$. Then, for any $\epsilon>0$, choose $\delta=\epsilon\sqrt{\pi}/8\mathrm{M}$. Then, for any $x>0$, the absolute value of the first term of the right member of (15) is less than $\epsilon/4$.

Next, if x is less than $\mu=\epsilon(\pi at)^{1/2}/4\mathrm{M}$, tne absolute value of the second term is less than $\epsilon/2$. Now $g(\tau)$ was assumed continuous on the left at $\tau=t$. Hence we can choose a positive number η such that $|g(t)-g(\tau)|<\epsilon/4$ whenever $0<t-\tau<\eta$. Hence, if $0<x<(4a\eta\delta^2)^{1/2}$, the third term of the right member of (15) will be less than $\epsilon/4$ in absolute value. Finally, $\left|\frac{\partial w}{\partial x}-g(t)\right|<\epsilon$ if $0<x<(4a\eta\delta^2)^{1/2}$, $0<x<\mu$, where η, μ, δ depend only on ϵ and t. Thus (12) is satisfied by (13).

It may be remarked here that left-handed continuity of $g(\tau)$ has not been required throughout the interval $[0, t]$. We have established the validity of (12) at any point t at which g is left-handed continuous.

---

* *Cf.* E. T. Whittaker and G. N. Watson, 'Modern Analysis,' 4th edition, § 4.2. Since $g(\tau)$ is bounded and measurable, the argument is obviously applicable to the Lebesgue integral considered here.

† *Cf.* E. C. Titchmarsh, 'The Theory of Functions,' (12.21) (1932).

Next, equation (11) is obviously satisfied. since the integrand in (13) is bounded for all $t \geqq 0$, $0 \leqq \tau \leqq t$, any fixed $x>0$.

To prove that (13) satisfies the differential equation (1) simply requires the admissibility of differentiation inside the integral sign, once with respect to t and twice with respect to x, and differentiation with respect to the upper limit. The formal result is

$$\frac{\partial w}{\partial t}=a\frac{\partial^2 w}{\partial x^2}=-(a/x)^{1/2}\int_0^t g(\tau)\{x^2/4a(t-\tau)^{5/2} \\ -1/2(t-\tau)^{3/2}\} \exp\left[-x^2/4a(t-\tau)\right] d\tau.$$

(See the reference to " Whittaker and Watson " given above.)

If the slab extends from $x=0$ to $x=-\infty$, with the same boundary and initial conditions, the second term in the right member of (8) becomes positive, and the first term is to be integrated from $-\infty$ to 0.

4. *Connexion with Classical Theory.*

First we obtain the surface temperature

$$\phi(t)=\lim w(x, t) \text{ as } x\to+0.$$

From (13)

$$\begin{aligned} w(x, t)+(a/\pi)^{1/2}\int_0^t g(\tau)(t-\tau)^{-1/2}d\tau \\ =-(a/\pi)^{1/2}\int_0^{t-\delta} g(\tau)(t-\tau)^{-1/2}\{-1+\exp\left[-x^2/4a(t-\tau)\right]\} d\tau \\ -(a/\pi)^{1/2}\int_{t-\delta}^t g(\tau)(t-\tau)^{-1/2}\exp\left[-x^2/4a(t-\tau)\right] d\tau \\ +(a/\pi)^{1/2}\int_{t-\delta}^t g(\tau)(t-\tau)^{-1/2} d\tau. \end{aligned}$$

The absolute value of the second term of the right member is smaller than $\mathrm{M}nd\delta(a/\pi)^{1/2}$, where N is the upper bound of $\tau^{1/2}\exp\left[-x^2/4a\tau\right]$ and M is the upper bound of $g(\tau)$ in $(0\leqq\tau\leqq t)$.

The third term is smaller, in absolute value, than $2\mathrm{M}(a\delta/\pi)^{1/2}$.

Hence, for any $\epsilon>0$, the sum of these two terms will be less, in absolute value, than $2\epsilon/3$, if we choose $\delta=(\pi/a)^{1/2}(\epsilon/6\mathrm{M}) \min(1, 1/\mathrm{N})$.

With this choice of δ we can choose η, so that

$$(t-\tau)^{-1/2}\left|-1+\exp\left[-x^2/4a(t-\tau)\right]\right|<\epsilon(\pi a)^{1/2}/3t\mathrm{M}$$

whenever $0<x<\eta$, since the function $\tau^{-1/2}\exp\left[-x^2/4a\tau\right]$ is a continuous function of x at $x=0$, uniformly in τ in $(\delta\leqq\tau\leqq t)$.

Thus we have shown that

$$w(x, t)\to-(a/\pi)^{1/2}\int_0^t g(t-\tau)\tau^{-1/2}d\tau \text{ as } x\to+0. \quad . \quad . \quad . \quad (16)$$

For a surface temperature $\phi(t)$, the solution is known * to be

$$u(x, t)=x/2(\pi a)^{1/2}\int_0^t \phi(t-\xi)\xi^{-3/2}\exp\,[-x^2/4a\xi]\,d\xi. \quad . \quad . \quad . \quad (17)$$

For a surface gradient $g(t)$ we have found, in (13), the solution

$$w(x, t)=-(a/\pi)^{1/2}\int_0^t g(t-\tau)\tau^{-1/2}\exp\,[-x^2/4a\tau]\,d\tau. \quad . \quad . \quad (18)$$

We shall show that either of these is formally obtainable from the other.

Using (18), we found in (16) that

$$\phi(t)=-(a/\pi)^{1/2}\int_0^t g(t-\tau)\tau^{-1/2}\,d\tau.$$

Taking Laplace transforms,

$$\Phi(s)=-\mathrm{G}(s)(a/s)^{1/2}, \quad \text{or} \quad \mathrm{G}(s)/s=-\Phi(s)/(as)^{1/2},$$

and the inverse Laplace transforms give

$$\int_0^t g(\tau)\,d\tau=-(\pi a)^{-1/2}\int_0^t \phi(t-\tau)\tau^{-1/2}\,d\tau.$$

Differentiating with respect to t,

$$g(t)=-\phi(0)/(\pi at)^{1/2}-(\pi a)^{-1/2}\int_0^t \phi'(t-\tau)\tau^{-1/2}\,d\tau.$$

Substituting in (18),

$$w(x, t)=\frac{1}{\pi}\int_0^t [\phi(0)(t-\tau)^{-1/2}+\int_0^{t-\tau}\phi'(t-\tau-\xi)\xi^{-1/2}\,d\xi]\tau^{-1/2}\exp\,[-x^2/4a\tau]\,d\tau.$$

Taking Laplace transforms,

$$\begin{aligned}\mathrm{W}(x, s)&=[\phi(0)s^{-1/2}+\{s\Phi(s)-\phi(0)\}s^{-1/2}]s^{-1/2}\exp\,[-x(s/a)^{1/2}]\\&=\Phi(s)\exp\,[-x(s/a)^{1/2}].\end{aligned}$$

Finally, the inverse Laplace transforms give

$$w(x, t)=x/2(\pi a)^{1/2}\int_0^t \phi(t-\xi)\xi^{-3/2}\exp\,[-x^2/4a\xi]\,d\xi,$$

which is equation (17).

To prove the converse, differentiate (17) with respect to x:

$$\frac{\partial u}{\partial x}=(\pi a)^{-1/2}\int_0^t \phi(t-\xi)\{1/2\xi^{3/2}-x^2/4a\xi^{5/2}\}\exp\,[-x^2/4a\xi]\,d\xi.$$

Integrating by parts, and setting $x=0$,

$$g(t)=-\phi(0)(\pi at)^{-1/2}-(a\pi)^{-1/2}\int_0^t \phi'(t-\xi)\xi^{-1/2}\,d\xi.$$

Taking Laplace transforms,

$$\mathrm{G}(s)=-\Phi(s)(s/a)^{1/2},$$

---

* Carslaw, *ibid.* pp. 46, 47.

and the inverse transforms give

$$\phi(t) = -(a/\pi)^{1/2}\int_0^t g(\tau)(t-\tau)^{-1/2}\,d\tau.$$

Substitution in (17) gives

$$u(x,t) = -x/2\pi\int_0^t \xi^{3/2}\exp\,[-x^2/4a\xi]\,d\xi\int_0^{t-\xi} g(\tau)(t-\xi-\tau)^{\ 1/2}\,d\tau.$$

Taking Laplace transforms,

$$\mathrm{U}(x,s) = -\mathrm{G}(s)(a/s)^{1/2}\exp\,[-x(s/a)^{1/2}],$$

and the inverse transform gives

$$u(x,t) = -(a/\pi)^{1/2}\int_0^t g(t-\tau)\tau^{-1/2}\exp\,[-x^2/4a\tau]\,d\tau,$$

which is equation (18).

Obviously, these formal proofs can easily be made rigorous.

5. *Special Case: $g(t)$ a Step Function.*

Let τ be a fixed positive number, and define $g(t)$ as follows:

$$g(t) = g,\ \text{a constant, for } 0 \leqq t \leqq \tau,$$
$$g(t) = 0,\ \text{for } t > \tau.$$

We have two cases to consider:

(*a*) $t \leqq \tau$. Equation (13) becomes

$$w(x,t) = -g(a/\pi)^{1/2}\int_0^t (t-\xi)^{-1/2}\exp\,[-x^2/4a(t-\xi)]\,d\xi.$$

Make the change of variable $\mu^2 = x^2/4a(t-\xi)$:

$$w(x,t) = -gx\pi^{-1/2}\int \xi^{\ 2}e^{-\xi^2}\,d\xi,\quad y = x/2(at)^{1/2}.$$

Integration by parts then gives

$$w(x,t) = gx - 2g(at/\pi)^{1/2}\exp\,[-x^2/4at] - gx\,\mathrm{erf}\,(y), \quad . \quad . \quad (19)$$

where
$$\mathrm{erf}\,(y) = 2\pi^{-1/2}\int_0^y e^{-\xi^2}\,d\xi,$$

and is tabulated, for example, in Peirce's Table of Integrals.

Using either (16) or (19), the surface temperature is found to be

$$w(0,t) = -2g(at/\pi)^{1/2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

(*b*) $t > \tau$. Equation (13) becomes

$$w(x,t) = -g(a/\pi)^{1/2}\int_0^\tau (t-\xi)^{-1/2}\exp\,[-x^2/4a(t-\xi)]\,d\xi.$$

Changing the variable and integrating by parts as before gives

$$w(x, t)=2g\pi^{-1/2}\{[a(t-\tau)]^{1/2}\exp[-y_1^2]-(at)^{1/2}\exp[-y_2^2]\}$$
$$+gx\{\operatorname{erf}(y_1)-\operatorname{erf}(y_2)\}, \quad . \quad (21)$$

where $y_1^2=x^2/4a(t-\tau), \quad y_2^2=x^2/4at.$

Using (16) or (21), the surface temperature is found to be

$$w(0, t)=2g\sqrt{\frac{a}{\pi}}(\sqrt{t-\tau}-\sqrt{t})=-2g\sqrt{\frac{at}{\pi}}\sum_{n=1}^{\infty}\frac{(2n-2)\,!}{2^{2n-1}n\,!\,(n-1)\,!}\left(\frac{\tau}{t}\right)^n \quad . \quad (22)$$

Obviously, $w(0, t)\rightarrow-2g\sqrt{\frac{a\tau}{\pi}}$ *as* $t\rightarrow\tau,$

in agreement with (20), and $w(0, t)\rightarrow 0$ *as* $t\rightarrow\infty$.

6. *Special Case : g(t) Periodic.*

Let $g(t)=\sin\omega t.$

From equation (16) the surface temperature is given by

$$w(0, t)=-(a/\pi)^{1/2}\int_0^t\sin\omega(t-\tau)\tau^{-1/2}d\tau.$$

Since $$\int_0^\infty x^{-1/2}\sin x\,dx=\int_0^\infty x^{-1/2}\cos x\,dx=\sqrt{\pi/2},$$

we can write

$$w(0, t)=-\sqrt{a/2\omega}(\sin\omega t-\cos\omega t)+\sqrt{a/\pi}\int_t^\infty\sin\omega(t-\tau)\tau^{-1/2}d\tau.$$

Now the integral can be written as a convergent, alternating series, each term being an integral over a range of length π/ω. Thus, the absolute value of the integral is less than the first term of the series which, in turn, is less than $\pi/\omega\sqrt{t}$.

Hence, finally,

$$w(0, t)=\sqrt{a/\omega}\cos(\omega t+\pi/4)+\mathrm{R}(t), \quad . \quad . \quad . \quad . \quad . \quad (23)$$

where $|\mathrm{R}(t)|<\sqrt{a\pi/t}/\omega$. For large values of t we may neglect R and use (23) as a new boundary condition. The resulting problem has been solved *.

The University of California.

---

* See Carslaw, *ibid.* pp. 47–50.

XXXI. *Effusion Phenomena in a Degenerate Bose-Einstein Gas.*

By D. V. Gogate, M.Sc., Ph.D. (London) and Y. V. Kathavate, B.Sc., Physics Department, Baroda College, Baroda (India) *.

[Received March 6, 1941.]

Introduction.

In recent years a number of papers have appeared on the investigation of effusion phenomena on the basis of Fermi-Dirac statistics, but the case of Bose-Einstein degeneracy has not been treated so far. In view of the revived interest (1) in Bose-Einstein statistics, it seems appropriate to place on record explicit expressions for the various properties of a Bose-Einstein gas, and in this paper we take up the study of effusion phenomena.

1. Let us consider an ideal gas enclosed in a vessel having a narrow orifice, and let the gas escape through this orifice into a vacuum outside. We shall assume that the mean free path of the particles is sufficiently great compared to the size of the orifice and also that the de Broglie wave-length associated with particles of average energy is small compared to its size, so that the diffraction effect † may be neglected.

Under these conditions we shall investigate the effusion of number, mass, and energy of the particles through the aperture.

---

* Communicated by the Authors.

† This condition has not be explicitely stated by the previous workers. If a plane wave is incident on a small circular aperture of radius " a," then, on emergence, the angle of spread is (for small values of θ), $\theta=\lambda/a$. This is a measure of the diffraction effect. Expressing the wave-length λ of the de Broglie waves, in terms of h, m, and v and substituting for v its value $(k\mathrm{T}/m)^{\frac{1}{2}}$, we get

$$\theta=\frac{h}{\sqrt{mk\mathrm{T}}}\cdot\frac{1}{a}.$$

If we consider, as an example, the case of helium gas at 1° K., we find

$$\frac{h}{\sqrt{mk\mathrm{T}}}=2{\cdot}76\times10^{-7},$$

and as the radius " a " of the aperture will be much greater than this value, the diffraction effect is negligible.

The general expression for the Bose-Einstein distribution law, taking into consideration the effect of relativity mechanics, is

$$n(\epsilon)d\epsilon = \frac{4\pi g}{c^3h^3}\,\frac{(\epsilon^2+2mc^2\epsilon)\,(\epsilon+mc^2)d\epsilon}{1/\mathrm{A}\,.\,e^{\epsilon/k\mathrm{T}}-1}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where $n(\epsilon)d\epsilon$ is the number of particles per unit volume having energies between ϵ and $\epsilon+d\epsilon$ and m, c, g, and h have their usual meaning. For non-degeneracy A is less than unity, and for degeneracy A is equal to unity.

2. *Effusion of particles.*—The number $\overline{\mathrm{N}}$ of particles, effusing per unit area per second, will be, as usual (2)

$$\overline{\mathrm{N}} = \int_0^\infty n(\epsilon)d\epsilon \int_0^{\pi/2} v\cos\theta\,.\,\frac{1}{2}\sin\theta d\theta = \int_0^\infty n(\epsilon)d\epsilon\,.\,\frac{v}{4} = \frac{1}{4}n\bar{v}, \quad . \quad (2)$$

where $\bar{v}$ is the mean velocity. Now since

$$v = c\left\{1-\frac{1}{(\epsilon/mc^2+1)^2}\right\}^{\frac{1}{2}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

we have

$$\overline{\mathrm{N}} = \pi gc\left(\frac{k\mathrm{T}}{ch}\right)^3 \int_0^\infty \frac{u^2+2uy}{1/\mathrm{A}\,.\,e^u-1}du, \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where

$$u = \frac{\epsilon}{k\mathrm{T}} \quad \text{and} \quad y = \frac{mc^2}{k\mathrm{T}}.$$

For the degenerate case ($\mathrm{A}=1$), equation (4) is easily integrated in terms of the Riemann Zeta function $\zeta(s)$, giving

$$\overline{\mathrm{N}} = \pi gc\left(\frac{k\mathrm{T}}{ch}\right)^3\left[2\zeta(3)+\frac{2mc^2}{k\mathrm{T}}\zeta(2)\right]. \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Equation (5) gives the general expression for the effusion of a degenerate Bose-Einstein gas into vacuum. We shall now consider two particular cases—(i.) non-relativistic, and (ii.) relativistic.

For the completely non-relativistic case ($y\to 8$), we have, from (5),

$$\overline{\mathrm{N}}_{\text{non-rel.}} = \pi gc\left(\frac{k\mathrm{T}}{ch}\right)^3 \cdot \frac{2mc^2}{k\mathrm{T}}\,\frac{\pi^2}{6} = mg\frac{\pi^3k^2}{3h^3}\mathrm{T}^2, \quad . \quad . \quad . \quad . \quad . \quad (6)$$

whereas for the completely relativistic case ($y\to 0$),

$$\overline{\mathrm{N}}_{\text{rel.}} = \pi g\frac{(k\mathrm{T})^3}{c^2h^3}\,.\,2\zeta(3) = 2{\cdot}404\frac{\pi g}{c^2}\,\frac{(k\mathrm{T})^3}{h^3}. \quad . \quad . \quad . \quad . \quad . \quad (7)$$

3. *Effusion of Mass.*—It can be easily shown that the mass effusing per unit area per second is given by

$$\overline{\mathrm{M}} = \frac{\pi g}{c^4h^3}(k\mathrm{T})^4[3\,!\,\zeta(4)+3y\,.\,2\,!\,\zeta(3)+2y^2\zeta(2)]. \quad . \quad . \quad . \quad (8)$$

In the completely non-relativistic case, (8) gives

$$\overline{\mathrm{M}}_{\text{non-rel.}}=\frac{2\pi g}{c^4h^3}(k\mathrm{T})^4y^2\,.\,\zeta(2)=\frac{\pi g}{3h^3}(\pi mk\mathrm{T})^2, \quad . \quad . \quad . \quad . \quad (9)$$

and in the completely relativistic case,

$$\overline{\mathrm{M}}_{\text{rel.}}=\frac{6\pi g}{c^4h^3}(k\mathrm{T})^4\frac{\pi^2}{90}=\frac{g}{15}\frac{\pi^3}{h^3}\left(\frac{k\mathrm{T}}{c}\right)^4. \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

4. *Effusion of Energy.*—The expression for the effusion of energy ⁽³⁾ is

$$\overline{\mathrm{E}}=\pi gc\left(\frac{k\mathrm{T}}{ch}\right)^3\int_0^\infty\frac{u^3+2u^2y}{e^u-1}k\mathrm{T}\,du. \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Hence the general expression for the energy effusing per second per unit area will be

$$\overline{\mathrm{E}}=\pi gc\frac{(k\mathrm{T})^4}{(ch)^3}[3\,!\,\zeta(4)+2y\,.\,2\,!\,\zeta(3)]. \quad . \quad . \quad . \quad . \quad (12)$$

The non-relativistic and relativistic expressions are given by

$$\overline{\mathrm{E}}_{\text{non-rel.}}=\frac{2\pi gc(k\mathrm{T})^4}{(ch)^3}\,.\,\frac{2mc^2}{k\mathrm{T}}\,.\,1{\cdot}202=4{\cdot}808\,\pi gm\left(\frac{k\mathrm{T}}{h}\right)^3 \quad . \quad . \quad (13)$$

and

$$\overline{\mathrm{E}}_{\text{rel.}}=\frac{6\pi g(k\mathrm{T})^4}{c^2h^3}\frac{\pi^2}{90}=\frac{g}{15}\frac{\pi^3(k\mathrm{T})^4}{c^2h^3}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

It will be noted that in degenerate Bose-Einstein statistics, the expressions for the effusion of number, mass, and energy depend only on temperature, whereas, in the non-degenerate case, they depend on concentration as well.

5. The formulæ (4) to (14) hold for the effusion of a degenerate gas into vacuum. Let us now calculate the net effusion from one chamber into another, the two being at different temperatures but containing the same gas. The formulæ for the effusion of number, mass, and energy in this case can be easily written down from the previous expressions by making use of the theorem proved below.

From formula (1) we can easily see that the transport of any quantity from chamber (1) to an evacuated chamber (2) will always be given by an equation of the form

$$\underset{(1\rightarrow 2)}{\mathrm{T}r}=\int_0^\infty\frac{\phi(\epsilon)d\epsilon}{1/\mathrm{A}\,.\,e^{\epsilon/k\mathrm{T}}+\beta}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

and the transport of this quantity when the chamber (2) already contains the same gas will be

$$\underset{(1\rightarrow 2)}{\mathrm{T}r}=\int_0^\infty\frac{\phi(\epsilon)}{1/\mathrm{A}\,.\,e^{\epsilon/k\mathrm{T}}+\beta}\left\{1-\beta\frac{n'(\epsilon')}{a(\epsilon')}\right\}d\epsilon, \quad . \quad . \quad . \quad (16)$$

where the probability factor in the brackets { } takes account of the already-occupiedness of the phase cells of the second chamber. $n'(\epsilon')d\epsilon'$ denotes the number of particles per unit volume in the second chamber having energies between ϵ' and $\epsilon'+d\epsilon'$, and $a(\epsilon')d\epsilon'$ denotes the corresponding number of phase cells. We assume that, when a particle is transported from one chamber to another, its energy remains the same, *i. e.*, the region of transport is field-free. Thus $\epsilon'=\epsilon$ and equation (16) reduces to

$$\underset{(1\rightarrow 2)}{\mathrm{T}r}=\int_0^\infty \frac{\phi(\epsilon)d\epsilon}{1/\mathrm{A}\,.\,e^{\epsilon/k\mathrm{T}}+\beta}-\beta\int_0^\infty \frac{\phi(\epsilon)}{a(\epsilon)}\,\frac{n'(\epsilon)}{1/\mathrm{A}\,.\,e^{\epsilon/k\mathrm{T}}+\beta}d\epsilon. \quad . \; . \quad (17)$$

The transport from chamber (2) to chamber (1) will similarly be

$$\underset{(2\rightarrow 1)}{\mathrm{T}r}=\int_0^\infty \frac{\phi(\epsilon)d\epsilon}{1/\mathrm{A}'e^{\epsilon/k\mathrm{T}'}+\beta}-\beta\int_0^\infty \frac{\phi(\epsilon)}{a(\epsilon)}\cdot\frac{n(\epsilon)d\epsilon}{1/\mathrm{A}'\,.\,e^{\epsilon/k\mathrm{T}'}+\beta}. \quad . \; . \quad (18)$$

The distribution law states that

$$n(\epsilon)\,(1/\mathrm{A}\,.\,e^{\epsilon/k\mathrm{T}}+\beta)=a(\epsilon), \quad . \; . \; . \; . \; . \; . \; . \quad (19)$$

and hence the second terms in the equations (17) and (18) are equal. Therefore the net transport from chamber (1) to chamber (2) will be, subtracting (18) from (17),

$$\underset{(1\rightarrow 2)}{\mathrm{T}r\ \mathrm{net}}=\int_0^\infty \frac{\phi(\epsilon)d\epsilon}{1/\mathrm{A}\,.\,e^{\epsilon/k\mathrm{T}}+\beta}-\int_0^\infty \frac{\phi(\epsilon)d\epsilon}{1/\mathrm{A}'\,.\,e^{\epsilon/k\mathrm{T}'}+\beta}, \quad . \; . \; . \quad (20)$$

that is, *the net transport has the same value as the difference between the transports from chamber* (1) *to vacuum and chamber* (2) *to vacuum.*

Hence the number of particles effusing from chamber (1) into chamber (2), in the non-relativistic and relativistic cases, will be

$$\overset{\text{non-rel.}}{\underset{(1\rightarrow 2)}{\mathrm{N}}}=\frac{mg\pi^3k^2}{3h^3}(\mathrm{T}_1{}^2-\mathrm{T}_2{}^2) \quad . \; . \; . \; . \; . \; . \; . \; . \quad (21$$

and

$$\overset{\text{rel.}}{\underset{(1\rightarrow 2)}{\mathrm{N}}}=\frac{2{\cdot}404\pi g}{c^2}\left(\frac{k}{h}\right)^3(\mathrm{T}_1{}^3-\mathrm{T}_2{}^3), \quad . \; . \; . \; . \; . \; . \quad (22)$$

where T_1 and T_2 denote the temperatures of chambers (1) and (2) respectively. Similarly, the expressions for the effusion of mass and energy from chamber (1) to chamber (2) can be immediately written down from the expressions previously given.

Summary.

In this paper the effusion phenomena in a degenerate Bose-Einstein gas are investigated, taking into account the effect of relativity mechanics. It is shown that, in degenerate Bose-Einstein statistics, the expressions

for the effusion of number, mass, and energy depend only on temperature whereas, in the non-degenerate case, they depend upon concentration as well. The effusion from one chamber into another, the two being at different temperatures but containing the same gas, is then calculated, and a simple and direct proof of the general theorem that " the net transport from chamber (1) to chamber (2) is equal to the difference between the transports from chamber (1) to vacuum and chamber (2) to vacuum " is also given.

References.

(1) F. London, Phys. Rev. liv. p. 947 (1938).
(2) B. N. Srivastava, Proc. Roy. Soc. A, clxvii. p. 517 (1938).
(3) D. V. Gogate, Phil. Mag. xxix. p. 87 (1940).

XXXII. *The Pressure Wave System in the Periodic Flow of Liquid in a Pipe.*

By R. S. Silver, M.A., B.Sc., Ph.D., Chief Physicist, G. & J. Weir, Ltd., Engineers, Cathcart, Glasgow, S. 4.*

[Received February 11, 1942.]

Further work has disclosed an analytical error in a previous paper on the above subject which appeared on p. 15 of the January 1942 issue of this Magazine. vol. xxxiii. The main theoretical development of the paper and all the matter of the appendix are unaffected.

With the limitations of accuracy in experimental work the arithmetical values derived from the erroneous formulæ had been accepted as satisfactory. Recently. however, as a result of applications requiring greater accuracy, an error was revealed. It affects only the discussion on p. 21 to the beginning of the appendix on p. 23, which should be replaced by the following :—

Now, by considering $z=\cos\theta+i\sin\theta=e^{i\theta}$ and using the expression for $\dfrac{1}{1-xz}$, it can be shown that

$$1+x\cos\alpha+x^2\cos 2\alpha+\ldots\text{etc.}=\frac{1-x\cos\alpha}{1-2x\cos\alpha+x^2}$$

and

$$x\sin\alpha+x^2\sin 2\alpha+\ldots\text{etc.}=\frac{x\sin\alpha}{1-2x\cos\alpha+x^2}.$$

$$\therefore\ p-p_{\text{mean}}=rq_mh\left\{[\sin(m\omega t-\beta)+\sin(m\omega t+\beta)]\left[\frac{1-x\cos\alpha}{1-2\,x\cos\alpha+x^2}\right]\right.$$
$$\left.-[\cos(m\omega t-\beta)+\cos(m\omega t+\beta)]\left[\frac{x\sin\alpha}{1-2x\cos\alpha+x^2}\right]-\sin(m\omega t+\beta)\right\}.$$

Now

$$\sin A+\sin B=2\sin\frac{A+B}{2}\cos\frac{A-B}{2};\ \cos A+\cos B=2\cos\frac{A+B}{2}\cos\frac{A-B}{2}.$$

$$\therefore\ p-p_{\text{mean}}=rq_mh\left\{2\left(\frac{1-x\cos\alpha}{1-2x\cos\alpha+x^2}\right)\sin m\omega t\cos\beta\right.$$
$$-\frac{2x\sin\alpha}{1-2x\cos\alpha+x^2}\cos m\omega t\cos\beta$$
$$\left.-\sin m\omega t\cos\beta-\cos m\omega t\sin\beta\right\}.$$

---

* Communicated by the Author

$$=rq_m h\left\{\sin m\omega t\left[\frac{2(1-x\cos\alpha)\cos\beta}{1-2x\cos\alpha+x^2}-\cos\beta\right]\right.$$
$$\left.-\cos m\omega t\left[\frac{2x\sin\alpha\cos\beta}{1-2x\cos\alpha+x^2}+\sin\beta\right]\right\}$$
$$=rq_m h\left\{\sin m\omega t\left[\frac{\cos\beta(1-x^2)}{1-2x\cos\alpha+x^2}\right]\right.$$
$$\left.-\cos m\omega t\left[\frac{2x\sin(\alpha-\beta)+\sin\beta(1+x^2)}{1-2x\cos\alpha+x^2}\right]\right\}.$$

Now write the coefficient of $\sin m\omega t$ as a_y and the coefficient of $\cos m\omega t$ as b_y.

$$\therefore\ p-p_{\text{mean}}=rq_m h[a_y\sin m\omega t-b_y\cos m\omega t]$$
$$=rq_m h\sqrt{a_y^2+b_y^2}\,[\sin(m\omega t+\theta)],\quad . \quad (9)$$

where θ is an angle whose tangent is b_y/a_y.

Equation (9) shows that, in these circumstances, the amplitude of the pressure variation at the point y is

$$rq_m h\sqrt{a_y^2+b_y^2},$$

and its frequency corresponds to that of the source.

Hence the maximum and minimum amplitudes are determined by the maximum and minimum amplitudes of $\sqrt{a_y^2+b_y^2}$, which must now be considered.

Expanding and simplifying

$$a_y^2+b_y^2$$
$$=\frac{(1-x^2)^2\cos^2\beta+4x^2\sin^2(\alpha-\beta)+\sin^2\beta(1+x^2)^2+4x(1+x^2)\sin(\alpha-\beta)\sin\beta}{(1-2x\cos\alpha+x^2)^2}.$$

We first consider the maximum and minimum values of this expression with respect to β, *i. e.*, we consider the positions in the pipe at which maxima and minima occur. There are turning points if

$$\frac{d}{d\beta}(a_y^2+b_y^2)=0.$$

By differentiating we find that this occurs when

$$\tan 2\beta=\tan\alpha.\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Equation (10) determines the positions in the pipe at which for any given value of α, *i. e.*, for any source frequency, the surge is greatest and least.

Now since 2β must be less than or equal to α the loci of maximum and minimum amplitudes will be given by

$$\beta=\frac{\alpha}{2}-\frac{S\pi}{2},$$

where S is any integer or zero.

Now $$\alpha=\frac{2m\omega L}{c} \quad\text{and}\quad \beta=\frac{m\omega y}{c},$$

so that the condition $\beta=\frac{\alpha}{2}$ corresponds to the end of the pipe distant from the source. Hence the maximum and minimum amplitudes will be alternately placed at intervals corresponding to $\frac{\pi}{2}$, *i. e.*, a distance between any two maxima will correspond to π. This length we may call the wave-length of the system, and if it is denoted by λ it will be given by

$$\frac{m\omega\lambda}{c}=\pi.$$

Therefore, if n is the frequency $\lambda=\frac{c}{2n}$.

It is now possible to consider the values of these maximum and minimum amplitudes. If we substitute

$$\beta=\frac{\alpha}{2}-\frac{S\pi}{2},$$

we obtain ultimately

$$a_y^2+b_y^2=\frac{1}{1-\dfrac{4x\cos^2\frac{\alpha}{2}}{(1+x)^2}},$$

when S is zero or an even integer, and

$$a_y^2+b_y^2=\frac{1}{1+\dfrac{4x\sin^2\frac{\alpha}{2}}{(1-x)^2}},$$

when S is an odd integer.

These latter equations indicate the amplitudes at alternate values of S so that, for a given value of x one of these will represent the maximum amplitude in the system, while the other will represent the minimum.

Now by differentiation, it is possible to determine the maximum and minimum values of these expressions, which themselves represent the maximum and minimum amplitudes occurring in the system.

Thus when $\cos\frac{\alpha}{2}=\pm 1$, we have a turning value, at positions corresponding to S an even integer, of

$$a_y^2+b_y^2=\frac{(1+x)^2}{(1-x)^2}.$$

The value at the intermediate positions of S is then

$$a_y^2+b_y^2=1.$$

This illustrates very clearly that when

$$\alpha=0,\ 2\pi,\ 4\pi,\ 6\pi,\ \text{etc.},$$

and if x is positive, the highest amplitude coefficient in the system is

$$\frac{1+x}{1-x},$$

while the lowest is 1. On the other hand, if x is negative under these conditions of α the highest amplitude coefficient in the system will be 1, while the lowest will be

$$\frac{1+x}{1-x}.$$

It will, therefore, be seen that if a length of line were so chosen that corresponding to a particular source frequency we had α equal to an even multiple of π, resonance conditions would be approached in the system if x were nearly equal to $+1$, and that in any case there would be a magnification of the original source disturbances as long as x was positive. If, however, x were negative, under these conditions of α there would be no magnification of the source disturbances. The practical importance of this equation is obvious.

Similarly, it can be shown that when x is negative and α is an odd multiple of π, resonance conditions will occur when x becomes nearly equal to -1, and will give multiplication of the source disturbances in any case, but that if α is an even multiple of π the resultant disturbance will not at any point exceed that originally due to the source.

(The original paper now resumes at Appendix I.)

XXXIII. *Notices respecting New Books.*

The Theory of Rate Processes. By SAMUEL GLASSTONE, KEITH J. LAIDLER, and HENRY EYRING. [Pp. x.+611.] (New York and London: McGraw Hill, 1941. Price 42*s*.)

MODERN physical chemistry is concerned with interpreting chemical reaction rates and equilibria in terms of the forces between atoms and molecules and in terms of molecular configuration. The two important tools which are used in the elucidation of kinetic and equilibrium problems are quantum mechanics and statistical mechanics.

During the last few years great strides have been made in this field, and the power and importance of these methods in dealing with problems of reaction rate and equilibria have been amply demonstrated. The actual use of these new methods has, however, been left in the hands of relatively few chemists, but the time has now come when these methods can and should be incorporated into the general structure of our chemical approach to new problems. Herein lies one of the main values of this book, in that it collects and discusses a very wide field of problems which have been dealt with. The treatment of these problems is now appearing for the first time in book form, set in the proper background of fundamental quantum mechanical and statistical theory. One of the early chapters on quantum mechanics, while naturally not going into the detailed fundamentals of the subject, gives a sufficient insight to enable one to use what we might term chemical quantum mechanics, namely, those approximations and generalizations which have been found most useful in the study of chemical problems. By means of the quantum mechanical approach to the theory of the reaction kinetics and activation energy it has been possible to define the configuration of the reacting system along the whole of the reaction path. This allows of the application of methods of statistical mechanics to reaction kinetics, and it is difficult to exaggerate the importance of the conception resulting from this. Here, again, this book gives a sufficient background of statistical theory to enable one to understand the application of this theory to rate processes.

Thus, having introduced the reader to the use of these two important tools, the authors then devote the rest of the book to the application of the methods to a very diverse field of chemical and physical rate processes. The application to homogeneous gas reactions was historically the first and has been worked out in the greatest detail. Naturally, therefore, in this book the chapter on homogeneous gas reactions is an important one and deals in detail with the application of the method to a large number of simple chemical reactions. In describing, however, the difference between such reactions as $H+H_2$, $D+D_2$, $H+HD$, etc., the theoretical methods are being subjected to a very severe test, and the results are not vastly different from those obtained by the much simpler method of the collision theory. The methods described in this book, however, have, in a large number of cases, succeeded where the collision theory has failed completely to give an adequate interpretation of the results of chemical reaction kinetics, and the success of the method is demonstrated under such applications as diene addition reactions, polymerization of butadiene, the combination of free radicals and unimolecular reactions. These methods, however, are not restricted to the simpler field of homogeneous gas reactions, and chapters are devoted to reactions involving excited electronic states and heterogeneous reactions. It is probable that the problems of reactions in solution, both physical and chemical reactions, have yielded the most fruitful field for the application of this method, which has been very successful in giving understanding of factors other than activation energy which influence the velocity of reaction and also in dealing with viscosity and diffusion processes as a branch of kinetics. In the whole of this work there are many problems which are as yet unsolved and problems to which the present approach is inadequate, but the book does not pretend to be a final statement on the subject, but rather an introduction to a method which will grow in power as it gains in exactitude.

In the opinion of the reviewer the value of this book in chemical science cannot be too highly praised. It may be argued that since the methods are frequently very approximate in their quantitative application that their value is small. This, however, is not the case, since the insight into the mechanism of chemical processes which these methods are able to give is of enormous conceptual importance, and for an autonomous advance of chemistry it is necessary that chemists should construct theoretical approaches to their problems based largely on the experimental foundations of their science.

M. G. E.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

XXXIV. *On the Dimensions of Physical Magnitudes.*

By HERBERT DINGLE, A.R.C.S., D.I.C., D.Sc.*

[Received February 23, 1942.]

I. INTRODUCTORY.

Several papers † have appeared recently on the subject of dimensions. They show wide disagreement, and on reading them one cannot help feeling that this arises from a lack of formulation (and, in at least some instances, lack of realization) of the fundamental principles underlying physical measurements and relations, which constitute the final Court of Appeal in such matters. This becomes less surprising when one discovers that the theory of dimensions has never been adequately based on general principles. Rules have often been laid down (*e. g.* by Clerk Maxwell and Brown), but they neither go sufficiently deep nor command general assent. Wilson, for instance, while in general following Clerk Maxwell, regards dimensions as characterizing physical quantities, while Clerk Maxwell speaks of dimensions of units. The main purpose of this paper is to relate the subject to more fundamental and universally acceptable ideas and thus to establish a set of principles to which the questions under discussion may be referred for settlement. A secondary purpose is to refer some of them.

II. DEFINITIONS AND POSTULATES.

I will first state a number of definitions and postulates which together form the principles to which I suggest all questions concerning dimensions should be referred. The statement is followed by comments and examples whose object it is to make the definitions and postulates clear and convincing. They contain nothing revolutionary, and each taken by itself will probably command immediate assent. It is the conclusions that inevitably follow from their combination that seem to have been overlooked.

---

* Communicated by the Author.

† G. B. Brown, Proc. Phys. Soc. lii. p. 577 (1940); liii. p. 418 (1941). W. E. Duncanson, Proc. Phys. Soc. liii. p. 432 (1941). W. Wilson, Phil. Mag. xxxiii. p. 26 (1942). G. D. Yarnold, Proc. Phys. Soc. liv. p. 46 (1942). W. E. Benham, Proc. Phys. Soc. liv. p. 121 (1942).

(1) *A physical quantity* may be taken as anything that can be measured by one or more strictly definable processes.

(2) A *measurement* of a physical quantity generally consists (and could, if desired, always consist) in principle of a numerical comparison of the quantity with an arbitrarily chosen *unit*; the result of the measurement, represented in physical equations by a symbol, will be called the *magnitude* of the corresponding physical quantity.

(3) Every magnitude appearing in a general physical equation represents the result of a measurement of a physical quantity by a unique, strictly specified process.

(4) Magnitudes are of two kinds—*fundamental* and *derived.* A fundamental magnitude is one whose value is unaltered by any change in the process of measurement, or in the chosen unit, of any physical quantity other than the one to which it refers. A derived magnitude is one whose value is in general altered by such a change.

(5) The number of fundamental magnitudes is arbitrary.

(6) A derived magnitude may be uniquely expressed in terms of those fundamental magnitudes, a change in the units or processes of measurement of which produces an alteration in its value.

(7) *Physical equations* are of two kinds—*definitions of derived magnitudes*, and *experimentally established physical relations.* Whether a given equation is of one or the other kind depends on the choice of fundamental magnitudes.

(8) When every magnitude occurring in a physical equation is reduced to fundamental magnitudes, every term in the equation consists of the same magnitudes raised to the same powers; *i. e.* the equation is homogeneous in each magnitude.

(9) The power to which each fundamental magnitude occurs in the reduced expression of a term in a physical equation is called the *dimension* of that fundamental magnitude in the corresponding term.

(10) Dimensions are characteristic of magnitudes, which are the results of measurements of physical quantities by strictly specified processes; they are not characteristic of physical quantities themselves.

(11) Physical equations may contain constants, as well as symbols representing magnitudes of measured quantities. In equations which define derived magnitudes, such constants always stand for definite, though arbitrarily chosen, numbers, and are dimensionless: in equations expressing experimentally established relations they are not arbitrary, their values being determined experimentally, and they are not dimensionless.

(12) The result of changing the unit in which any physical quantity is measured, the process of measurement being otherwise unchanged, is to multiply the magnitude of the quantity by a numerical factor. In view of the dimensional homogeneity of physical equations stated in postulate (8), such equations are therefore valid whatever the sizes of the units chosen.

Comments. (1) *A physical quantity may be taken as anything that can be measured by one or more strictly definable processes.* It is very difficult to define a physical quantity, but fortunately it is unnecessary since there is no disagreement here. Mass, time, elasticity, heat capacity, magnetic permeability, etc., are immediately acceptable as physical quantities. New physical quantities arise, and old ones vanish, from time to time; thus entropy, unknown to Newton, is now generally recognized as one, while the caloric of Carnot has passed out of existence. Whether a physical quantity is something existing objectively in nature and discovered, or waiting to be discovered, by us, or is a concept which we voluntarily form for convenience in rationalizing our experience, is a question still more fundamental than those now under consideration, and will not be touched upon. Whether discovered or invented, we may for our present purpose regard physical quantities as having an objective existence, with the essential property of being measurable.

They are, in general, measurable in various ways. Thus, length may be measured by a metre rod; by triangulation; by the time a specified moving body takes to traverse it, etc. Mass may be measured by inertia; by gravitational effects; by the number and kind of elementary particles in the body whose mass is required, etc. Velocity may be measured by the space covered in a stated time; by the Doppler effect; in terms of an arbitrarily chosen unit velocity *, etc.

The results of measurements of the same quantity by different methods may or may not be compatible. Thus, measurements of a length by a metre rod and by triangulation may be different, in which case, if we wish to use both methods, we must state which has final authority. If we choose the metre rod, then we say that space is non-euclidean; if we choose triangulation, we say the metre rod has expanded or contracted. To get unambiguous results it is clear that for every physical quantity we must choose a particular process of measurement which we regard as canonical. We must define the process completely and unequivocally, and any measurement made by an alternative process must, if necessary, be corrected in order to obtain the true measurement. The choice of the canonical process is entirely a matter of convention ; it should obviously be made by international agreement. How far this ideal is realized, and what degree of permanence should belong to the conventional choice, will be briefly discussed in Section III.

---

* A mechanism for such a process is suggested by Bridgman, 'The Logic of Modern Physics' (Macmillan, 1928), p. 99.

The fact that alternative processes of measurement of the same physical quantity do not inevitably give the same result has led thinkers of positivist schools to discard the idea of a physical quantity altogether and to grant meaning only to the process and result of measurement in each case. On this view measurements by a metre rod and by triangulation, respectively, represent two independent entities, no more fundamentally related to one another than either is to a measurement by a thermometer, for example : their agreement in most cases is then to be regarded as an empirical law of nature, like the agreement between E and CR as measures of the electrical state of a circuit. While this view is doubtless formally correct, we shall not adopt it since it introduces unnecessary elaboration and, moreover, tends to obscure relationships which it is the object of physics to bring to light. We shall still speak of physical quantities measurable by different methods, but since we shall never introduce a physical quantity apart from the measurement of it and shall, further, regard each process of measurement as distinct from every other, we shall preserve the rigour of the true positivist view without suffering from its inconveniences.

(2) *A measurement of a physical quantity generally consists* (*and could, if desired, always consist*) *in principle of a numerical comparison of the quantity with an arbitrarily chosen unit ; the result of the measurement, represented in physical equations by a symbol, will be called the magnitude of the corresponding physical quantity.* That a unit is involved in most measurements is suggested by the fact that the results of such measurements are never precisely stated unless they each contain a number followed by a word or words. Thus, a length of 3 is meaningless : it must be 3 feet, or 3 yards, or 3 something else. The word indicates the unit and the number indicates how many of such units are contained in the quantity measured. Sometimes the unit is a composite one ; thus, we may speak of an acceleration of 10 cm. per second per minute. In this case the unit of acceleration is 1 cm. per second per minute, and the acceleration in question contains 10 such units.

The only exception to this in current physics is in the measurement of angles. Here the angle to be measured is made to form part of a sector of a circle, and the angle is measured by the ratio of the lengths of the arc and the radius. This ratio is the same whatever unit of length is chosen, provided it is the same unit for both arc and radius ; it is, therefore, unnecessary for the number to be followed by a word. Nevertheless, we do use the word "radian," apparently indicating a unit, because we may, if we wish, measure the arc and radius in different units of length and so measure the angle in different "units" (*e. g.* in "degrees").

Such a procedure is possible because, in the measurement of length, no attention is paid to direction; the statement that a body is 6 feet long tells us nothing about the direction in which it lies. This neglect of vector properties is doubtless a defect in the current practice of measurement.

Although angles are the only "dimensionless" magnitudes generally recognized as occurring in physical equations, they are not the only possible ones, nor are they dimensionless "by nature"; it is a matter of convention. A "standard angle" could have been preserved at Paris beside the "standard metre," in which case angles, like lengths, would have been measured in terms of an arbitrary unit and would have had dimensions of their own. On the other hand, velocity, for instance, could have been measured in terms of the relative Doppler displacement, $\frac{d\lambda}{\lambda}$. The process of measurement would have been to stand in the line of motion with a spectroscope and to measure this ratio. Velocity would then have been "dimensionless," but could have been expressed in terms of different "units" by measuring $d\lambda$ and λ in different units. This possibility again arises from the neglect of direction in measurement. $d\lambda$ varies with direction of observation, and it is necessary to state that the measurement must be made in the direction of motion in order to describe the process of measurement precisely.

The only point we wish to emphasize with regard to this postulate, however, is that although, in current practice, the only magnitudes which are not expressed as multiples of an arbitrarily chosen unit are those of angles, this does not belong to the nature of things, but is the result of convention.

(3) *Every magnitude appearing in a general physical equation represents the result of a measurement of a physical quantity by a unique, strictly specified process.* The chief importance of this postulate lies in the last five words. The magnitude represents not a physical quantity but the result of the measurement of a physical quantity, and that measurement is not a measurement by *any possible* method, but by some particular method whose specification is understood. Thus, the perfect gas equation, $PV = R\theta$, is true when temperature is measured on the absolute scale, but not when it is measured on the centigrade scale; and if we measure the mass of a body by the number of atoms multiplied by the

average atomic weight, we shall not verify the equation, $m=m_0\left(1-\frac{v^2}{c^2}\right)^{-\frac{1}{2}}$. It is true that an equation often remains valid, at least over a limited range, when the implied processes of measurement are changed to others, but that must be regarded as, in a sense, accidental. Fundamentally the equation represents relations between magnitudes obtained by uniquely specified processes of measurement.

(4) *Magnitudes are of two kinds—fundamental and derived. A fundamental magnitude is one whose value is unaltered by any change in the process of measurement, or in the chosen unit, of any physical quantity other than the one to which it refers. A derived magnitude is one whose value is in general altered by such a change.* An example of a fundamental magnitude is a length measured in terms of the standard metre, or a time measured in terms of the mean solar second. No change in the method or unit of measurement of any other physical quantity makes any difference to the result obtained in such a measurement. Examples of derived magnitudes are velocity, measured in units of space covered per unit of time; density measured in units of mass contained in a unit of volume; stress measured in units of force per unit of area, etc. Any of these magnitudes is changed by a change in the measurement of one of the component magnitudes. Thus, if the unit of mass is changed from the gram to the kilogram, the magnitude of the density of a body is thereby reduced one thousandfold.

(5) *The number of fundamental magnitudes is arbitrary.* It is probable that at least one magnitude must be fundamental, but there is no upper limit. In mechanics it has been customary to define the processes of measurement of length, time and mass so that the magnitudes of those quantities are fundamental; and the processes of measurement of all other quantities so that the resulting magnitudes are derived. This, however, is by no means necessary. Time, for instance, could have been measured by defining equal times as those in which a specified moving body covered equal lengths. In that case a time magnitude would have been derived, since it might have been changed by a change in the process of measuring length. On the other hand, velocity, which, when defined as $\frac{ds}{dt}$, is a derived magnitude, would have been fundamental if it had been defined as $\frac{d\lambda}{\lambda}$, as already suggested. There seems no reason why, given sufficient ingenuity, every magnitude should not be definable in a fundamental way.

(6) *A derived magnitude may be uniquely expressed in terms of those fundamental magnitudes, a change in the units or processes of measurement*

of which produces an alteration in its value. By postulate (4), any derived magnitude, a, may be expressed by the equation

$$a=f(b, c, \dots), \qquad (1)$$

where $b, c, \dots$ are the magnitudes a change in which affects it. Any of the magnitudes, $b, c, \dots,$ which are not fundamental can similarly be expressed as a function of other magnitudes, and the process can therefore be continued until a is expressed as a function of fundamental magnitudes only. That the expression is unique follows from the condition that every magnitude represents the result of a uniquely specified process of measurement.

(7) *Physical equations are of two kinds—definitions of derived magnitudes, and experimentally established physical relations. Whether a given equation is of one or the other kind depends on the choice of fundamental magnitudes.* Consider the equation

$$v=\frac{ds}{dt}, \qquad (2)$$

where v, s and t represent velocity, space and time, respectively. According to current practice, this is taken as a definition of the derived magnitude, velocity, and is therefore automatically true. The equation

$$v=\alpha\frac{d\lambda}{\lambda}, \qquad (3)$$

where α is a constant, is then, for moderate velocities, an experimentally established result. We may, however, as we have seen, take

$$v=\frac{d\lambda}{\lambda} \qquad (4)$$

as a definition of the measurement of velocity. In that case,

$$v=\beta\frac{ds}{dt}, \qquad (5)$$

where β is a constant, is a relation to be tested by experiment. It is purely conventional which choice we make ; there is nothing in nature to decide one way or the other.

The existence of these two types of equation leads, of course, to intermediate types, obtained by combining two or more "pure" equations. Thus, if we define velocity by equation (2) and acceleration, a, by the equation

$$a=\frac{dv}{dt}, \qquad (6)$$

we can deduce

$$a=\frac{d^2s}{dt^2}, \qquad (7)$$

and this is neither a primary definition of a derived magnitude nor an experimentally established relation. It is, however, logically merely another form of the definition of a, and so, together with all equations resulting from a combination of definitions, may be regarded as itself a definition without leading us into error. On the other hand, equations obtained by combining experimental relations with definitions or with other experimental relations—such as, for instance,

$$a=\gamma\frac{M}{r^2}, \qquad (8)$$

obtained by combining

$$F=ma, \qquad (9)$$

the definition of force, with

$$F=\gamma\frac{Mm}{r^2}, \qquad (10)$$

the experimentally found law of gravitation—are themselves experimentally established relations, since they do not define any magnitudes and their validity rests on experimental grounds. It is therefore unnecessary to complicate our considerations by contemplating more than two independent types of equation.

(8) *When every magnitude occurring in a physical equation is reduced to fundamental magnitudes, every term in the equation consists of the same magnitudes raised to the same powers* ; i. e., *the equation is homogeneous in each magnitude.* For equations representing definitions of derived magnitudes this is a truism. For other equations it expresses the necessity that a physical relation shall have a meaning. For example, the equation

$$48 \text{ inches}+3 \text{ hours}=4 \text{ feet}+180 \text{ minutes} \qquad (11)$$

is mathematically true but physically meaningless, because the process of adding magnitudes of length to those of time is not a physically significant one. It is obvious that any equation is similarly meaningless in which every term is not expressible in terms of the same magnitudes.

This is the basis of the principle of dimensional homogeneity which plays such a useful part in the derivation of physical laws. In order to apply the principle it is necessary to reduce each term to fundamental magnitudes, because only then is it expressed uniquely. Thus, the equation

$$v=\sqrt{\frac{\epsilon}{\rho}} \qquad (12)$$

for the velocity of longitudinal waves is not homogeneous as it stands. Each term must be reduced to fundamental magnitudes before the principle is seen to be satisfied.

(9) *The power to which each fundamental magnitude occurs in the reduced expression of a term in a physical equation is called the dimension of that fundamental magnitude in the corresponding term.* This is simply a definition, agreeing with generally acknowledged, but not always followed, convention.

(10) *Dimensions are characteristic of magnitudes, which are the results of measurements of physical quantities by strictly specified processes ; they are not characteristic of physical quantities themselves.* This results immediately from definition (9). It follows that we cannot properly speak of the dimensions of temperature, for example, but only of the dimensions of the magnitude of temperature measured by a strictly specified process. Thus, if we measure the temperature of a body by the reading of an air thermometer, say, we obtain a fundamental magnitude, θ. If, however, we measure it by the rate at which the body radiates energy, we obtain a derived magnitude which we find to be proportional to θ^4. It is obvious that the dimensions differ in the two cases. It is obvious also that, since the magnitude of the same physical quantity may be either fundamental or derived, according to definition, there can be no unique dimensions of a physical quantity.

(11) *Physical equations may contain constants, as well as symbols representing magnitudes of measured quantities. In equations which define derived magnitudes. such constants always stand for definite, though arbitrarily chosen, numbers. and are dimensionless : in equations expressing experimentally established relations they are not arbitrary, their values being determined experimentally, and they are not dimensionless.* This may be illustrated by equations (2) to (5). (2) is a definition of a derived magnitude, velocity. Now a method of measurement is not completely defined unless the unit of measurement is stated. As we have said, a length of 3 is meaningless ; it must be 3 units of some kind. Hence, if we take an equation as a definition, there must be no *unspecified* constant occurring in it. We may, it is true, choose the definition so as to imply any unit we please—we may, for instance, define velocity as $\frac{ds}{dt}$ or $10\frac{ds}{dt}$ or $\frac{1}{15}\frac{ds}{dt}$, and so on—but if we define it as $A\frac{ds}{dt}$, it is not defined at all unless A is some definitely specified number. Such an indefinite equation may serve to indicate that the *physical quantity*, velocity, is conceived as depending somehow on the *physical quantities*, space and time, but it tells us nothing about the relation of the *magnitude*, velocity, to the *magnitudes*, space and time. Consequently no unspecified constant can occur in an equation *defining a derived magnitude.*

On the other hand, constants such as α and β in (3) and (5), which occur in *experimentally established physical relations*, are beyond our control. The magnitudes occurring in the equation having been previously defined, such constants have their own values, not chosen by us, which observation will reveal. Their values will, of course, depend on the units in which the various magnitudes are measured, but, those units having been chosen, they are absolutely determined. Further, since an experimentally established relation is in general a relation between magnitudes having different dimensions (*e. g.* equations (3) and (5)), the constants which appear must have dimensions in order to satisfy the principle of dimensional homogeneity.

This distinction is important for our subsequent work. It will be convenient, therefore, to distinguish the two kinds of constant in our notation, and we shall do so by writing arbitrary, purely numerical constants, occurring in definitions, in capital Latin letters, and non-arbitrary constants, occurring in experimentally established relations, in small Greek letters.

The difference of character of these constants will be further discussed in Section V. in connexion with Benham's claim that equation (9), expressing a form of Newton's second law of motion, should contain an arbitrary dimensional constant, N.

(12) *The result of changing the unit in which any physical quantity is measured, the process of measurement being otherwise unchanged, is to multiply the magnitude of the quantity by a numerical factor. In view of the dimensional homogeneity of physical equations stated in postulate* (8), *such equations are therefore valid whatever the sizes of the units chosen.* The first part of this postulate is self-evident, in view of postulate (2). The second part follows because the numerical factors introduced by the change of units must be the same in every term of the equation, and can therefore be cancelled out.

It may be noted that this postulate illustrates the fact, to which attention was directed under postulate (2), that there are in current physics no units of angle. In the formula

$$t=2\pi\sqrt{\frac{l}{g}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

for the time of swing of a simple pendulum, for instance, we may, if we wish, speak of 2π as 2π radians, but if we decide to express angles in degrees, the coefficient does not become $2\pi\times 57{\cdot}3$.

III. AMBIGUITY IN MODERN PHYSICS OF CHOICE OF FUNDAMENTAL MAGNITUDES.

The previous section sets out certain conditions which must be fulfilled by physical theory. A large degree of liberty remains, especially in the choice of methods of measurement of physical quantities, and particularly in the selection of fundamental magnitudes. The next step is to see how physics uses this liberty. What are the fundamental magnitudes adopted, and how are the others derived from them ? The answer to this question reveals the source of much of the confusion which exists in this subject. The nominal magnitudes of physics are not always the actual ones ; they are not always equivalent to the actual ones ; and physicists generally are not aware that there is anything irregular.

Confining ourselves to mechanics, for simplicity, we find three traditional fundamental magnitudes—length, time and mass. Length is defined in terms of a standard rod (unit, 1 cm.), time in terms of the rotating earth (unit, 1 mean solar second), and mass in terms of a standard piece of matter and the conservation laws (unit, 1 gm.) *. Every length, mass and time magnitude is professedly expressed in terms of one of these arbitrary standards, and the method of measurement is generally understood. In recent years, however, many alternative proposals have been put forward—not legitimately, as candidates for conventional adoption in place of the existing conventions, but as though no convention now existed or as though magnitudes were fundamental by nature and not by convention. Thus Milne † proposes to recognize only one fundamental magnitude—that of time—and to derive all other magnitudes therefrom. Eddington ‡ protests that Milne, in thus overriding convention, is simply calling black white, but himself overrides it still more violently by asserting, without apology, that all standards of measurement have a quantum specification. Brown proposes to make mass and either length or time derived magnitudes, but for some reason retains both length and time in his expression of other derived magnitudes in what purport to be fundamental terms.

It is clear that all this is very unsatisfactory. It is not incumbent upon us to retain a convention that may have outgrown its desirability, simply because it was once established by general agreement, but it is necessary, first, that if it is superseded the change also shall be made by general agreement, and secondly, that until it is superseded the original convention shall be universally observed. As things are at present, one often does not know what is meant when certain magnitudes are mentioned. This would be bad enough if only formal expressions were at stake, but actually matters are much worse ; it is the laws which

---

* For an account of the principle of application, see, for example, my ‘ Special Theory of Relativity ’ (Methuen, 1940), pp. 62–3.

† ‘ Relativity, Gravitation and World Structure ’ (Oxford Univ. Press, 1935) and many subsequent papers.

‡ ‘ The Philosophy of Physical Science ’ (Camb. Univ. Press, 1939), Chapter 5.

our equations express that have become ambiguous, and the ambiguity is not realized.

Two examples must suffice. The point to be illustrated is this. The conventional definition of a magnitude has been changed to one which seems equivalent, and the fact that a change has been made has, therefore, been ignored. Further progress has shown that the two definitions are *not* equivalent. Contradictions have thus been introduced into physics, the source of which is unrealized, and a mass of futile argumentation on irrelevant points has thus been generated. The examples we choose are the magnitudes of temperature and time. Electrical and magnetic magnitudes, discussed in Section V, illustrate a slightly different misfortune, arising from the absence of *any* precise definition of the processes of measurement.

The Measurement of Temperature. The temperature of a body was originally defined in terms of the reading of a thermometer immersed in the body. The unit was the *degree*, which was one-hundredth of the change (of length, resistance, etc.) of some standard body when raised from the temperature of melting ice to that of boiling water. Variations in the kind of change and the choice of standard body gave a variety of scales, but this caused no confusion since the scale used was always stated, and the readings were, in fact, reduced to those of an ideal " perfect gas " thermometer. The measured temperature, θ, according to this definition was clearly a fundamental magnitude, for no change in the measurement of any other magnitude could make any difference to it.

Progress in physical theory, supported by all the experimental evidence available, then indicated that the temperature of a body as so defined was proportional to the mean kinetic energy (kinetic energy per unit mass) of the molecules of the body. Accordingly an equation was obtained,

$$\theta=\alpha\frac{E}{M}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

where α is a constant. Reduced to fundamental magnitudes, and expressed dimensionally, this becomes

$$[\theta]=[\alpha L^2T^{-2}], \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

whence the dimensions of α were given by

$$[\alpha]=[\theta L^{-2}T^2]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

All this is quite straightforward.

But then equation (14) became tacitly adopted as a *definition* of temperature, α, of course, having a specific numerical value assigned to it, and, therefore, becoming dimensionless. The dimensions of θ

thereupon became $[L^2T^{-2}]$; *i. e.*, temperature became a derived instead of a fundamental magnitude. Discussions on dimensions thus became richly endowed with meaningless verbosity, prevented from collapse only by unawareness that, first, two conflicting specifications of temperature measurement were current, and, secondly, dimensions are characteristic of such specifications and not of the physical quantity measured.

The next stage was the bringing to light of circumstances in which the conflicting specifications gave not only different dimensions for temperature, but also different *numerical results*. In interstellar matter, for example, the kinetic energy per unit mass corresponds to a temperature of some thousands of degrees, but there is every reason to believe that a thermometer placed there would read about 3° absolute *. Speaking in terms of the kinetic theory, the molecules move with high speed, but their collisions with the thermometer would be so rare that the energy imparted to it would be radiated away as fast as it was received. It therefore becomes still more imperative that the word " temperature " shall have a unique conventional meaning assigned to it. So long as it has not, physicists will be unable to talk a common language.

The Measurement of Time. The second example is the measurement of time, which I have discussed elsewhere † more fully in connexion with the interpretation of the special theory of relativity; a brief statement will be sufficient here.

When length is measured in terms of a standard rod, and time in terms of the rotation of the Earth, Newton's First Law of Motion becomes a hypothesis to be tested. One test (indirect, but valid) is the comparison of the observed with the calculated tracks of ancient eclipses ‡. A discrepancy is found, which means that Newton's law is inaccurate.

Astronomers, however, do not draw this conclusion; they say that the Earth is slowing down, while Newton's law remains true. But this means that the rotating Earth is abandoned as the standard of time measurement, and the scale implied by Newton's First Law is substituted for it. Equal times are then, *by definition*, those in which an undisturbed body moves over equal lengths, and the approximate uniformity of rotation of the Earth becomes a fact of observation. A still further change was made when Einstein substituted a beam of light for an undisturbed body; the " postulate of the constancy of the velocity of light" is, in effect, a definition of the scale for measuring time. Tacitly adopting this scale, Eddington can say, "my personal conclusion is that there is no more danger that the velocity of light . . . . will change with time than that the circumference-diameter ratio π will change with time." * Such a conclusion is possible only if light defines the time-scale, but by stating it as a "personal" conclusion, Eddington gives the impression that it is conceivably false. The active existence of two incompatible time scales in physics is thus clearly seen, and the reluctance of physicists to admit the fact, let alone provide a remedy, is amply illustrated in the correspondence following the articles in 'Nature' referred to above.

---

* See, for example, Eddington, 'The Internal Constitution of the Stars' (Camb. Univ. Press, 1926), Chapter 13.

† 'Nature,' cxliv. p. 888 (1939); cxlvi. p. 391 (1940).

‡ See Oppolzer, 'Canon der Finsternisse.'

Let us now see how the *dimensions* of time are affected by this duality. If time is measured in terms of the rotating Earth, it is a fundamental magnitude, and its dimensions are simply [T]. If, on the other hand, it is measured in terms of the space covered by a moving body or by light, it is a derived magnitude, for a change in the measurement of length would make a change in the value of a time magnitude. The equation (choosing light instead of an undisturbed body, for example, and choosing 1 cm. as the distance defining the unit of time) is

$$t=\frac{1}{3.10^{10}}l, \qquad . \ . \ . \ . \ . \ . \ . \ . \ . \ . \quad (17)$$

whence t must have the dimensions [L]. Hence again we have incompatible definitions yielding different dimensions, and until we decide how time is to be measured we cannot assign dimensions to any magnitude derived from time.

IV. THE PRACTICAL ASPECT OF THE CHOICE OF FUNDAMENTAL MAGNITUDES.

It would seem, then, that the situation calls for a general revision of physical measurements and a common agreement on how they are to be specified and which are to be fundamental. Let it be repeated that a re-affirmation of the traditional fundamental magnitudes is not necessarily to be desired. The aim of physics is to establish as comprehensive a system of relationships as possible between measurements of physical quantities, and it is obviously desirable to choose measurements which will make the subsequent work as simple as possible. Any choice is legitimate. We might, for example, measure all lengths relatively to a "growing" tree, but inasmuch as we should then find most bodies contracting we should have introduced a needless source of complication into our measurements. We have found something analogous to this in the discovery that the Earth is slowing down, relatively to a beam of light, and that most natural processes are more simply related to the time system defined by light than to that defined by the rotation of the Earth. There is, therefore, strong reason for abandoning the fundamental character of time measurement and deriving it from the measurement of length. (It need scarcely be said that this would make no difference whatever to the process of determining time at observatories for practical purposes.) What is most desirable, however, is that the change shall be made openly and by common agreement, and adhered to consistently in all theoretical physics.

---

* 'The Philosophy of Physical Science,' p. 78.

It is no part of our purpose to suggest a scheme of measurement, but it may be of interest to consider whether a large or a small number of fundamental magnitudes is desirable. Discussion of this point is by no means new *, but we are perhaps now in a better position than formerly to understand the issues involved. Let us take two extreme cases—first, that in which every magnitude is fundamental; and second, that in which all magnitudes except one are derived. We shall confine ourselves to mechanical magnitudes, for simplicity.

In taking all magnitudes as fundamental, we are assuming that a measurement of each is possible which does not depend on the method of measuring any other. This is certainly true of many—*e. g.* velocity, density, area, ... , which can be measured without separate measurements of space and time, mass and volume, lengths in different directions, ... and it is probable, and for purposes of illustration permissible to assume, that it is true of all. In that case, every physical relation would be an experimental discovery, and since the units in which the separate magnitudes were measured would be quite independent of one another, a dimensional constant would appear in every physical relation. Thus, the discovery that the velocity of a body was proportional to the space covered by it in unit time would be written,

$$v=\alpha\frac{l}{t}, \quad \ldots\ldots\ldots \quad (18)$$

and the dimensions of α would accordingly be

$$[\alpha]=[\mathrm{VTL}^{-1}]. \quad \ldots\ldots\ldots \quad (19)$$

Similarly, the kinematical law, $v=at$ (a=acceleration), would become

$$v=\beta at, \quad \ldots\ldots\ldots \quad (20)$$

and the dimensions of β would be

$$[\beta]=[\mathrm{VA}^{-1}\mathrm{T}^{-1}]. \quad \ldots\ldots\ldots \quad (21)$$

---

* See, for example, Tolman, Phys. Rev. ix. p. 237 (1917).

Furthermore, we might discover that

$$f = \gamma a m, \qquad (22)$$

where f and m are force and mass, respectively. This would give

$$[\gamma] = [\mathrm{FA}^{-1}\mathrm{M}^{-1}]. \qquad (23)$$

It is clear that every physical relation would have to be discovered by experiment and would introduce a fresh "constant of nature" whose dimensions would be determinable in terms of the measured magnitudes.

Restricting ourselves, however, for purposes of illustration to the relations already stated (a restriction which, of course, simply means setting ourselves at a stage at which experiment has carried us only so far), we can use (19), (21) and (23) to express V, A and F in terms of M, L, T and the constants, thus obtaining

$$\left.\begin{aligned} [\mathrm{V}] &= [\alpha \mathrm{LT}^{-1}], \\ [\mathrm{A}] &= [\alpha\beta^{-1}\mathrm{LT}^{-2}], \\ [\mathrm{F}] &= [\alpha\beta^{-1}\gamma \mathrm{MLT}^{-2}]. \end{aligned}\right\} \qquad (24)$$

(There would, of course, be no special reason for expressing all measured magnitudes in terms of the three, M, L and T, but we do so here for comparison with the actual situation in which these magnitudes are taken as fundamental.)

Let us now consider the application of the principle of dimensional homogeneity in these circumstances, taking as our example the time of swing of a simple pendulum, the amplitude being small. Assuming, as usual, that the time of swing, t, depends only on the mass, m, and weight, w, of the bob, and the length, l, of the string, we put

$$t = \mathrm{A} m^a w^b l^c, \qquad (25)$$

where A is a numeric. The dimensional equation is then

$$[\mathrm{T}] = [\mathrm{M}]^a [\alpha\beta^{-1}\gamma \mathrm{MLT}^{-2}]^b [\mathrm{L}]^c, \qquad (26)$$

which gives us 6 equations for 3 unknown quantities. They are easily seen to be compatible only if $a = c = \frac{1}{2}$, $b = -\frac{1}{2}$ and $\alpha\beta^{-1}\gamma$ is dimensionless, which gives us the familiar relation

$$t = \mathrm{A}\sqrt{\frac{ml}{w}}, \qquad (27)$$

together with the relation (from the last equation of (24)),

$$[\mathrm{F}] = [\mathrm{MLT}^{-2}]. \qquad (28)$$

Further, since $\alpha\gamma$ has the same dimensions as β, we have, from (21),

$$[\alpha\gamma] = [\mathrm{VA}^{-1}\mathrm{T}^{-1}]. \qquad (29)$$

This, we might imagine, would give us some insight into the nature of the constants α and γ. One cannot help being reminded of the similar relation

$$[\mu\kappa]=[\mathrm{L}^{-\frac{1}{2}}\mathrm{T}^{\frac{1}{2}}]. \qquad (30)$$

It is evident that in these circumstances the principle of dimensional homogeneity is a most fruitful source of discovery. A single application has yielded us (27), (28) and (29), whereas in the ordinary case it yields only (27). On the other hand, we do not in the ordinary case need (28) and (29); the former we establish by definition and the latter we regard as telling us nothing about nature but merely showing relations between constants introduced by our method of procedure. The principle is more successful in making discoveries because there are more discoveries to be made, and for all its success it leaves us worse off than when we make v, a and f derived magnitudes, because we are still encumbered by incompletely elucidated "constants of nature."

Let us now go to the other extreme, choosing only one fundamental magnitude, which we will take to be length. We then define equal times as those in which a moving body (*e. g.* light) covers equal lengths, and write

$$t=\mathrm{A}l, \qquad (31)$$

where A is now a dimensionless number, since we must give it a specific value before we have defined our measurement of time. For simplicity, therefore, we will take it to be 1, and we will similarly take all such numbers, which merely specify units, to be 1. We then make the definitions

$$\left.\begin{aligned} v&=\frac{dl}{dt}, \\ a&=\frac{dv}{dt}, \end{aligned}\right\} \qquad (32)$$

and obtain the dimensional relations

$$\left.\begin{aligned} [\mathrm{L}]&=[\mathrm{L}], \\ [\mathrm{T}]&=[\mathrm{L}], \\ [\mathrm{V}]&=[1], \\ [\mathrm{A}]&=[\mathrm{L}^{-1}]. \end{aligned}\right\} \qquad (33)$$

Mass is already expressed as length in general relativity, but since the relation is not expressible in simple terms as a definition of m we will use Newtonian language and define mass and force in terms of the two equations

$$f=ma=\frac{m^2}{l^2}, \qquad (34)$$

giving the dimensional relations

$$\left.\begin{array}{l}[\mathrm{M}]=[\mathrm{L}]\\ [\mathrm{F}]=[1]\end{array}\right\} . \qquad (35)$$

With these results let us again consider the time of swing of a simple pendulum. Equation (25) now gives

$$[\mathrm{L}]=[\mathrm{L}^a 1^b \mathrm{L}^c], \qquad (36)$$

whence

$$a+c=1, \qquad (37)$$

and b may have any value. This tells us very little, and it is clear that by deriving as many magnitudes as possible from a single fundamental one, we sacrifice most of the value of the principle of dimensional homogeneity as an instrument of research. It is obvious that since, for example, l, t and m all have the same dimensions, the principle must fail to distinguish between them.

The net result of these considerations, then, is that if we wish to give the greatest scope of usefulness to the principle of dimensional homogeneity, we must choose sufficient fundamental magnitudes to prevent two or more magnitudes from having the same dimensions, but not so many that indeterminable " constants of nature " are forced into our equations. In mechanics the traditional choice of three—M, L and T—comes near to satisfying these conditions, and is possibly the nearest practicable approach to the ideal. It does, however, involve some ambiguities ; for instance, it gives kinetic energy and torque the same dimensions.

V. APPLICATIONS TO PARTICULAR PROBLEMS.

A few of the questions at issue in recent papers may now be briefly considered. We choose three—the significance of " Newton's Constant," N, in the equation $f=\mathrm{N}ma$ * ; the dimensions of temperature ; and the character of the electromagnetic quantities, κ and μ. The first two have been discussed mainly by Benham, and the last by Brown, Yarnold and others. We shall not consider thermodynamic magnitudes, since we are concerned chiefly with principles and not the extent of their application, and there is nothing in thermodynamics that cannot be treated in the same manner as the magnitudes here discussed.

---

* Actually, Newton's Second Law should be expressed as $f \propto \frac{d}{dt}(mv)$. This corresponds more closely to his own statement and to our present knowledge that m is a function of v, and, therefore, of t when v is not constant. For our present purpose, however, the form given, which is that adopted by both Brown and Benham, is permissible.

"*Newton's Constant.*" Benham claims that the equation connecting force, mass and acceleration should be written

$$f = \mathrm{N}ma, \qquad \qquad (38)$$

where N is a dimensional constant. It follows at once, however, from the comments on postulate (11) that the validity of this claim depends entirely on whether the equation is a definition or an experimentally established relation between magnitudes previously defined. It is only in the latter case that N can have dimensions. The choice is, in the last resort, arbitrary, but it is the universal practice to take this equation as a definition of force. It if is not, then we have no right to use the symbol f in it, unless we have an antecedent definition to tell us what it means, and Benham proposes none. We must therefore assume that he regards f as defined by (38). He is then definitely wrong in claiming that N has dimensions.

It is of interest to compare the equations

$$f = ma \qquad \qquad (39)$$

and

$$f = \gamma \frac{mm'}{r^2}, \qquad \qquad (40)$$

the latter being Newton's law of gravitation. Since the first defines the measurement of force it contains no unspecified constant. As it stands it implies that the unit of force is the force which produces unit acceleration in unit mass. We are not bound to adopt this unit. We might, for instance, choose as the unit the force which produces unit acceleration in 8 units of mass. We must then substitute the equation $f = \frac{1}{8} ma$. But we cannot write $f = \eta\, ma$, where η is something of unknown dimensional character, without ceasing to define f at all. In equation (40), on the other hand, a dimensional constant γ must necessarily appear, for, the dimensions of f, m and r having been given by definition, those of γ must be such as to satisfy the principle of dimensional homogeneity.

It might be worth while to point out one consequence of granting Benham's claim. We should then have no ground for excluding a dimensional constant from every equation defining a derived magnitude. Thus we should write velocity $= \alpha \frac{ds}{dt}$; acceleration, $= \beta \frac{dv}{dt}$; area $= \delta l^2$; volume $= \epsilon l^3$; density $= \zeta \frac{m}{\mathrm{V}}$; and so on. We should then have our equations choked by a multitude of dimensional constants, each hinting at some dark secret of nature, while remaining essentially indeterminable. That way madness lies.

The Nature of Temperature. Benham's further belief that he has been examining the nature of temperature has already been discussed in the remarks on postulate (10) and in Section III., and calls for little further comment. Since the dimensions of temperature have been shown to vary with the method of measurement, and one method is fundamentally as legitimate as another, it follows that the dimensions give us information about the method of measurement, but not about the physical quantity, temperature. We shall see the same point illustrated in the following discussion of electrical and magnetic magnitudes.

Electrical and Magnetic Magnitudes. In no branch of physics is there so much confusion on fundamental points as in electricity and magnetism. The origin of this is only partly in the essential difficulties of the subject; it is mainly in the unsystematic way in which the subject is approached. As Duncanson wisely says: "If there is doubt concerning the dimensions of a quantity this means, not that there is some hidden mystery as to the complete nature of that quantity, but merely that the quantity concerned has not been unambiguously defined." * This, of course, follows immediately from postulates (10) and (3), that dimensions refer to magnitudes and not to the essential nature of physical quantities, and that magnitudes represent the results of measurement by strictly specified processes. If Duncanson had only paid equal attention to postulates (7) and (11), he would have penetrated much further into the heart of the problems he has discussed.

It follows, however, from his quoted remark, that the existing doubt concerning the dimensions of electrical and magnetic magnitudes indicates that these magnitudes have not been explicitly and unambiguously defined. They are unambiguously defined *implicitly*, or we should not agree on our measurements of them, as we do. Our problem, then, is to state precisely the definitions that are implied in practice.

The ambiguity in current expressions arises in the following way. It is stated by most of the recent writers that there is no equation between electrical and magnetic magnitudes which contains only one such magnitude. The most familiar example is the following, which is taken as a definition of electric charge:

$$\frac{q_1 q_2}{\kappa r^2} = f. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

But this is no definition of q unless κ, r and f are already defined, and it is equally no definition of κ unless q, r and f are already defined. Furthermore, it is not an experimentally established relation unless all the quantities are defined except one, which appears as the dimensional constant necessary to all such relations. It therefore belongs to neither of the two types of equation allowed by postulate (7) until we have independently defined κ or q.

---

* *Loc. cit.* p. 435.

I suggest the following as expressing the process of measurement of q implied in current practice. Choose any arbitrary standard charge, place the charge to be measured at a fixed position with respect to it *, the whole system being in a vacuum, and measure the force between the charges; the magnitude of the unknown charge is then taken as proportional to this force. We then have

$$q = \mathrm{N}f, \qquad \ldots \ldots \ldots \qquad (42)$$

where N is some definite number which we can specify as we wish, and the dimensional equation is

$$[\mathrm{Q}] = [\mathrm{F}], \qquad \ldots \ldots \ldots \qquad (43)$$

i. e. the dimensions of charge are those of force, because a change in the measurement of force is the only change that can affect the measurement of charge. This is a process which in principle can be carried out, and its results are precisely equivalent to the measures of charge which we actually use.

Now suppose two charges are measured in this way, and found to be q_1 and q_2. We then find, *by experiment*, that when they are separated by a distance r, they exert a force on one another given by (41), where κ is a constant whose numerical value in each case the experiment reveals; it is analogous to the reciprocal of the gravitational constant. The dimensional equation is

$$[\mathrm{F}^2\kappa^{-1}\mathrm{L}^{-2}] = [\mathrm{F}], \qquad \ldots \ldots \ldots \qquad (44)$$

so that

$$[\kappa] = [\mathrm{FL}^{-2}]. \qquad \ldots \ldots \ldots \qquad (45)$$

We have thus obtained the dimensions of q and κ, each unambiguously expressed, without in any way altering the meanings of those magnitudes as they are used in existing theory, but simply by describing existing practice in a systematic way. If, as usual, we choose M, L and T as fundamental units, the dimensions are

$$\left.\begin{aligned} [\mathrm{Q}] &= [\mathrm{MLT}^{-2}], \\ [\kappa] &= [\mathrm{ML}^{-1}\mathrm{T}^{-2}]. \end{aligned}\right\} \qquad \ldots \ldots \ldots \qquad (46)$$

Other definitions of q are, of course, possible. We might have measured q in terms of the heating effect of a current, for example, in which case we should have obtained quite different dimensions. There is no contradiction in this, for the dimensions characterize the method of measurement, and on changing the method we should expect to change the dimensions. The method of measurement suggested, however, is the one which corresponds to the familiar electrostatic system of units, and leads most directly to dimensions of κ. From what has been said it should be clear that it is quite irrelevant to look to (46) for a clue to the "physical nature" of κ.

---

* Not at *so many units of space* from it, for then the distance would vary with the unit of space chosen, but at a *fixed position* with respect to it. The standard charge and the charge to be measured might, for example, be placed one at each end of the standard metre at Paris, whether or not the standard metre is used as a unit of length. The adoption of such a standard framework is precisely analogous to the adoption of a standard kilogram or standard space-measuring rod in the measurement of mass or length.

It might be noted that (46) includes no fractional indices. This would be regarded by some writers (*e. g.* Duncanson) as a merit, but we cannot claim any fundamental significance for it. Fractional indices are not necessarily meaningless. It would have been legitimate, for example, to define charge as proportional, not to the measured force, but, say, to the $\frac{13}{89}$th power of the measured force. In that case fractional indices would have been rampant, and would have been perfectly significant. Or we might have chosen area instead of length as a fundamental magnitude, in which case the square root of an area magnitude would have had the obvious meaning of a length magnitude.

The root of the objection to fractional indices is again the idea that dimensions characterize physical quantities instead of the arbitrarily adopted method of measuring them. In case the falsity of this is not yet evident, we might give one further example. The usual expression for the dimensions of acceleration is $[LT^{-2}]$; *i. e.* acceleration may be measured in cm. per sq. sec. But what is a square second? There is only one dimension of time. Regarded as a characteristic of a physical quantity, $[T^{-2}]$ is just as meaningless as $[M^{\frac{1}{2}}]$, say, but as characteristics of chosen methods of measurement, both may be perfectly significant.

Now let us turn to magnetic magnitudes. We can, if we wish, make some magnetic magnitude a fundamental one, or measure magnetic, like electrical, magnitudes directly in mechanical terms, but at the present stage of progress it seems more satisfactory to derive such magnitudes from electrical ones. Again we shall choose definitions which are equivalent to those universally implied, but here unambiguously expressed.

We start by defining the measurement of a magnetic field. Take a compass needle constrained by some force (not necessarily magnetic) to lie in a certain direction. To measure the magnetic field at a point, place the direction of the constraining force perpendicular to that of the field, and observe the deflexion of the needle—θ, say. We then define the intensity of the magnetic field as proportional to $\tan\theta$.

Our equation is then

$$H = N \tan\theta \qquad \text{. . . . . (47)}$$

where H represents the field and N is some definite number which we can specify as we wish ; and the dimensional equation is

$$[\mathrm{H}]=[1], \quad \dots\dots\dots\dots \quad (48)$$

i. e. magnetic field is dimensionless.

Now we find by experiment that a magnetic field exists at the centre of a circular coil carrying a current. We define the current as the charge passing any point of the circuit in unit time, and we measure it by indirect processes which are equivalent to measuring charge and time We can also measure r, the radius of the coil, and H, the field created at the centre, and we find by experiment the relation

$$\mathrm{H}=\frac{\mu q}{rt}, \quad \dots\dots\dots\dots \quad (49)$$

where μ is constant. The dimensional equation is therefore

$$[1]=[\mu \mathrm{F} \mathrm{L}^{-1}\mathrm{T}^{-1}], \quad \dots\dots\dots \quad (50)$$

whence

$$[\mu]=[\mathrm{F}^{-1}\mathrm{L}\mathrm{T}]=[\mathrm{M}^{-1}\mathrm{T}^{3}]. \quad \dots\dots\dots \quad (51)$$

Now μ, defined in this way, is equivalent to permeability. Hence the dimensions of permeability, according to definitions expressing explicitly our existing practice, are $[\mathrm{M}^{-1}\mathrm{T}^{3}]$.

It is easily seen that

$$[\mu\kappa]=[\mathrm{L}^{-1}\mathrm{T}], \quad \dots\dots\dots\dots \quad (52)$$

i. e. $\frac{1}{\mu\kappa}$ has the dimensions of velocity. As ordinarily expressed, this quantity has the dimensions of the square of a velocity. That is because in (41), mistakenly regarded as a definition of charge although it defines nothing at all, q appears to the second power ; in other words, charge is treated as proportional to the square root of the force instead of to the force itself. (There is an equivalent procedure with regard to μ.) Another important difference is that in the ordinary expression the individual dimensions of κ and μ are unknown, whereas here they are known. That they are those of stress and the reciprocal of (stress × velocity) has no fundamental importance.

The above derivation of κ and μ puts them on precisely the same footing as, say, the gravitational constant, γ. These are all constants appearing in equations expressing experimental relations ; they are not direct measurements of " physical quantities " like mass or velocity, which can be measured by a variety of methods. There is nothing to tell us, for instance, that κ and μ will vary from one material to another ; that must be found by experiment. We find, in fact, that they do vary, whereas γ does not. It is this fact, and not the dimensions which we find them to possess, that justifies us in expecting that they will be expressible in terms of the physical properties of substances.

If we calculate the capacity of a condenser (*i. e.* the charge necessary to create unit difference of potential between the plates) from the law of force, we find it to be proportional to κ. The ratio of capacities of two condensers, having different dielectrics but otherwise identical, is therefore $\frac{\kappa_1}{\kappa_2}$. This is obviously dimensionless. It is not, however, κ, and Brown is therefore incorrect in saying that κ is dimensionless because it is the ratio of two similar measurements. If, however, κ_2 refers to a condenser with a vacuum as dielectric, and we choose units of q, f and r such that $\kappa_2=1$, then κ_1 becomes *numerically* equal to the dimensionless ratio $\frac{\kappa_1}{\kappa_2}$; but it by no means follows that κ_1 is dimensionless. The principle of dimensional homogeneity applied to (41) gives the same dimensions, whatever units are chosen, by postulate (12).

It is, of course, open to Brown to restrict his attention in the first place to K, the ratio of the capacities of the two condensers referred to above; this quantity will be dimensionless because its value will not be altered by a change in the unit of measurement of any physical quantity. But if he then wishes to calculate the capacity of one of the condensers, he will need a law of force, and he will not be able to adopt (41) without introducing another constant, κ, which he has, of course, no right to identify with K. Further work will then reveal to him that K is equal to the ratio of the κ's for the dielectrics in the two condensers, but he will not be able to make κ dimensionless.

Finally, the question may be raised whether the expression of the dimensions of electrical and magnetic magnitudes in terms of those of purely mechanical magnitudes is a realization of the 19th century aspiration to reduce all physical phenomena to mechanics. The answer is that it is no more than might have been obviously possible from the beginning. All electrical and magnetic phenomena, as observed, are simply motions. We do not observe electric charges or magnetic poles or electromagnetic fields; we simply observe that bodies move in peculiar ways. Consequently all electrical and magnetic magnitudes must be expressible in terms of the magnitudes which serve to describe motion, *i. e.* mechanical magnitudes. We need not, of course, so express them unless we wish. We can make new fundamental magnitudes for them, as we should do, for example, if we measured charge as a multiple of the electronic charge. The point once more is that dimensions characterize magnitudes, which are the results of measurements, and if our measurements are measurements of mechanical quantities the dimensions will be dimensions of mechanical magnitudes.

XXXV. *An Investigation of the Time-Lag in an Electromagnetic Clutch.*

By F. Horner, M.Sc., Electrotechnics Department,
Manchester University *.

[Received January 30, 1942.]

Introduction.

During recent tests on an electromagnetic clutch it was found that considerable time elapsed between switching off the exciting current and the separation of the plates. With a separating force of 90 lb. the time-lag was of the order of a second, which was too high a value when quick release was required. A short experimental investigation was therefore carried out to decide whether the time-lag could be attributed solely to eddy currents in the clutch, and to provide a basis for a more satisfactory design of clutch.

Details of the Clutch.

The clutch is illustrated in fig. 1. It consisted of a ring of steel of approximately circular cross-section, cut into two parts along a plane normal to the ring axis. One part, subsequently called the magnet, contained the magnetizing coil, while the other, the keeper, served to complete the magnetic path round the coil. The trough in which the coil was embedded divided the contact surface into two parts, each of

Fig. 1.

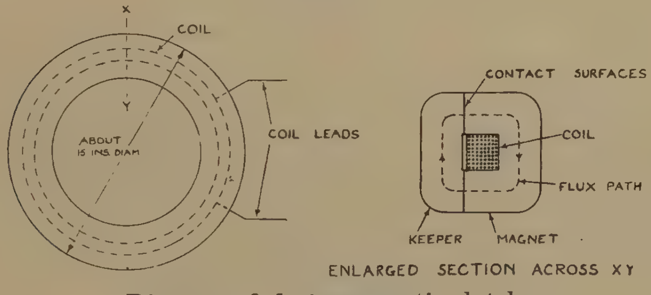


Diagram of electromagnetic clutch.

---

* Communicated by Prof. Willis Jackson.

area 201 sq. cm., and connexions to the coil were made through insulated leads passing through holes in the steel.

The friction surfaces were smooth machine-finished, but not optically flat. It was to be expected, therefore, that a small, irregular air-gap existed between these faces when the keeper was in position, but apart from this a closed path of steel existed round the energizing coil.

The coil contained 96 turns and the normal exciting current was about $4\frac{1}{2}$ amps.

Flux-current Curve and Hysteresis Loop of the Clutch.

The variation of flux round the coil with energizing current was studied with the aid of a fluxmeter. The magnetizing winding did not completely fill the slot provided and the additional space was used to house a search coil. A fluxmeter, connected to this, measured changes in the total flux linking with the coil, the contact surfaces being pressed together by the weight of the keeper.

Fluxmeter deflexions were read for reversal of current in the coil, the reversal being carried out several times for each current value to stabilize conditions. The hysteresis loop was traced by a similar method and the results of both these tests are shown in fig. 2.

The sharp rise in flux with current at about 1 amp. is noteworthy, since no corresponding change was noticeable in a B–H curve taken on a ring specimen of the same steel. It is due, presumably, to changes in the effective thickness of the small irregular air-gap between the contact faces at certain degrees of magnetization.

Force Required to Separate the Clutch Faces.

After the two parts of the clutch have been magnetically coupled and the current switched off, separation of the faces does not occur if the mechanical separating force is insufficient to overcome the attraction due to residual magnetism. Assuming that all the flux is confined to the steel and evenly distributed over the contact surfaces, the relationship between attractive force P and total flux ϕ is

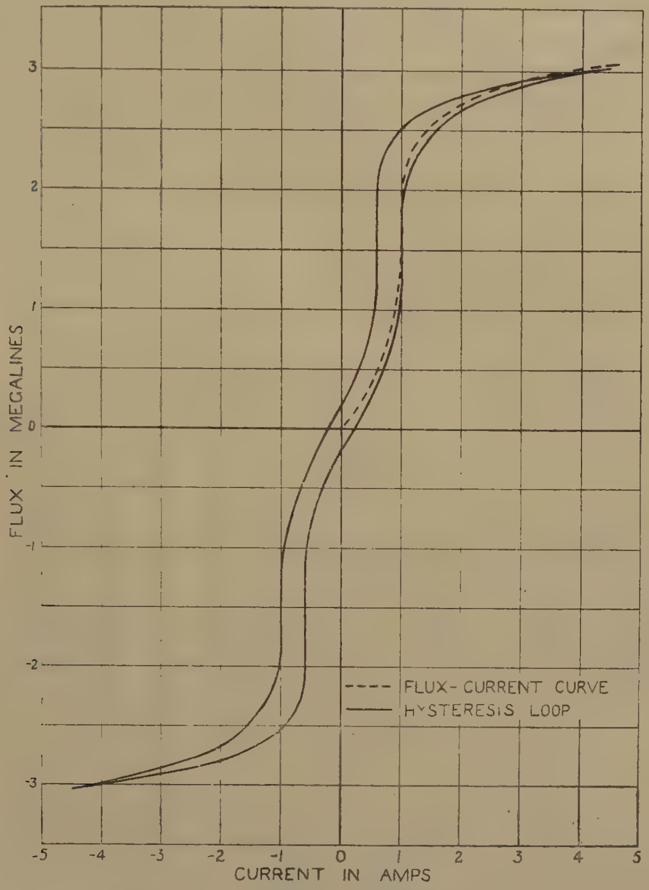
$$P=891\phi^2, \quad \ldots\ldots\ldots \quad (1)$$

P being expressed in lbs. if ϕ is in megalines.

According to the hysteresis loop the residual flux following normal excitation was 0·19 megalines, corresponding to an attractive force of 32 lb., and this should be the minimum force required for separation.

To measure the minimum separating force the magnet was placed in a horizontal position with the contact surface downwards. The keeper was fixed to the underside with normal excitation and the current broken. By adding weights to a platform attached to the keeper the minimum separating force, including the weight of the keeper, was found to be 29 lb.

Fig. 2.



Flux-current curve and hysteresis loop.

The discrepancy between the measured and predicted values is probably due to two causes :—

(1) Reduction of flux density in the gap through fringing and flux in the space surrounding the steel.

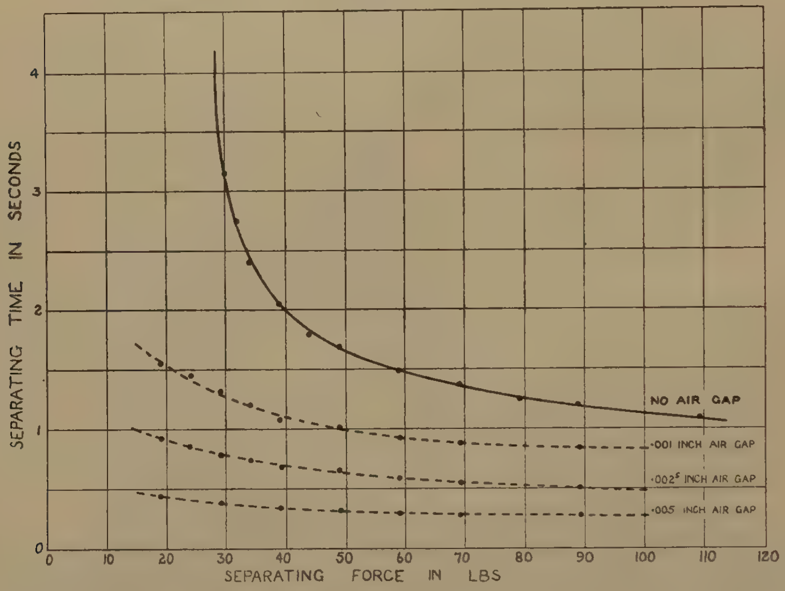
(2) The hysteresis loop was taken with the faces pressed together. Under the influence of a separating force the air-gap would be greater and the residual flux correspondingly less.

On the other hand, any irregularities in the air-gap would cause concentrations of flux which would increase the attractive force due to residual flux. This effect was evidently less than the combination of (1) and (2).

Variation of Separating Time with Applied Separating Force.

The experimental procedure described in the last section was used to find the dependence of separating time on applied separating force, initial magnetization always being normal. The clutch was engaged, a separating force applied and the time for the keeper to fall after breaking the current measured with a stop-watch to the nearest 1/10th second. In this and all subsequent measurements of time, recorded values represent the means of several readings. By thus averaging the errors it was considered that results could be obtained to greater accuracy than the 0·1 second of individual measurements, and it was considered that the personal error also lay well within this limit.

Fig. 3.



The results are plotted in fig. 3 (full curve) and are consistent with the following simple explanation.

After the breaking of the magnetizing current the total flux ϕ_t can be considered to consist of two parts, the residual flux ϕ_r, assumed constant, and the remainder $(\phi_t - \phi_r)$ which falls gradually to zero.

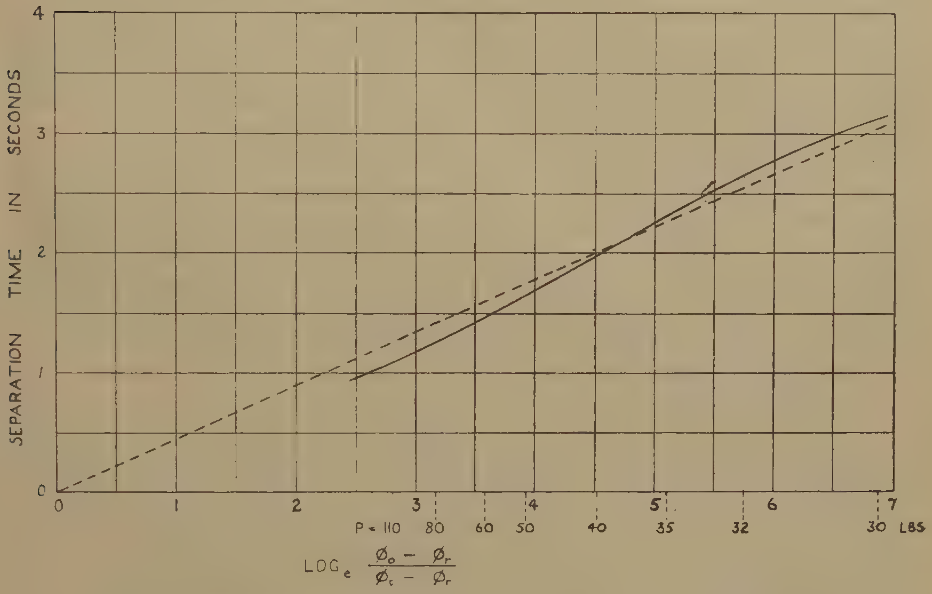
The changing flux cannot induce a current in the magnetizing coil since the circuit is broken, but there remains a complete conducting circuit in the steel with which the flux links and in which eddy currents are induced. This eddy current path is effectively a coil with a time-constant T seconds.

If T were strictly constant the flux linking with this path at time t seconds after breaking the circuit would be given by

$$\phi_t = (\phi_0 - \phi_r)e^{-t/\mathrm{T}} + \phi_r, \quad \ldots \ldots \ldots \quad (2)$$

where ϕ_0 is the initial flux.

Fig. 4.



If separation occurs when the flux has fallen to ϕ_t, t is the separation time. Hence if t is plotted against $\log\left\{\frac{\phi_0-\phi_r}{\phi_t-\phi_r}\right\}$ a straight line graph should result. Here ϕ_0 corresponds to the initial exciting current and ϕ_t to the known separating force. The residual flux ϕ_r can be derived approximately from the hysteresis loop, but probably with greater accuracy from the minimum separating force ($891\phi_r^2$). Fig. 4 (full curve) has been derived from fig. 3 in this way.

Some deviation from linearity is to be expected since T is not constant, as assumed, owing to the change of permeability with magnetization. In fact T varies considerably with magnetization, as can be seen from fig. 2, and greater deviation from linearity than is shown in fig. 4 was to be expected. It is considered, however, that the results support the view that the time-lag was due to eddy currents in the steel. It is evident from fig. 3 that very large separating forces would be required to reduce the time-lag to a small fraction of a second.

Variation of Separation Time with Resistance of the Eddy-Current Path.

Further confirmation of the preceding theory was sought through a study of the effects of varying the resistance of the eddy-current path. The clutch was closed and the current broken, the magnetizing coil being simultaneously closed through an external resistance. The time of separation was measured for several values of this resistance, the separating force remaining constant at 54 lb. With the closing of the coil there were two eddy current paths in parallel, one through the coil circuit of known resistance R_c, and one through the steel, the effective resistance (R_s) of which was unknown. It seems reasonable to assume that R_s was sensibly constant, while R_c was varied externally.

The time-constant (T) of the equivalent circuit formed by the two paths in parallel can be expressed in the form $\frac{L}{R}$, where L is the total inductance and R is the resultant of R_s and R_c in parallel. Changes in R_c do not affect the inductance, which may therefore be taken as constant. The separation time should be proportional to T and hence to $\frac{1}{R}$, so it can be expressed as

$$t \propto \left\{\frac{1}{R_c}+\frac{1}{R_s}\right\}. \qquad (3)$$

Since R_s is constant, a plot of t against $\frac{1}{R_c}$ should give a straight line graph. The recorded results have been plotted in this form in fig. 5, in which the departure of individual points from the mean straight line are within the experimental accuracy.

Effect of Air-Gap Thickness.

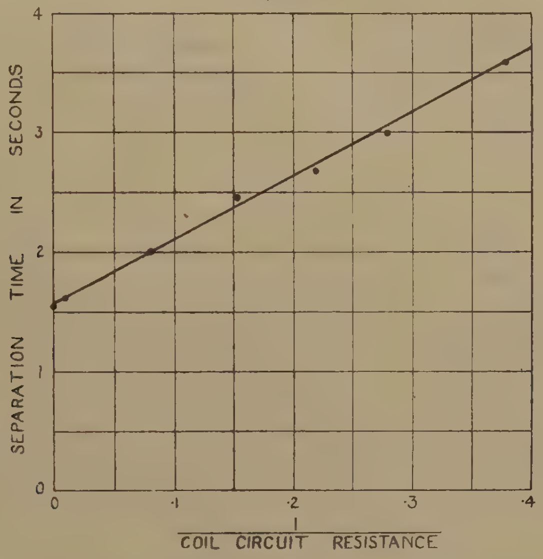
Since there was always a small air-gap between the faces of the clutch it was of interest to study the behaviour of the clutch with different air-gap thickness. This was done by interposing sheets of paper between the faces, and the effect of the gap thickness on the separation time is shown in fig. 3 (dotted curves). The reduction in time-lag with increasing gap thickness was due in part to the lower flux density (the magnetization current being the same), but flux-current curves showed that the 0·001-inch gap reduced the maximum flux by only 7 per cent. The reduction in inductance, therefore, accounted for most of the change. The marked effect of the 0·001-inch gap on the separation time indicates that normally the effective air-gap was extremely small.

Methods of Reducing Time-Lag.

The electromagnetic clutch has many varied applications and, should it be necessary to reduce time-lag, the method must be chosen with due regard to such items as available space, electric power supply, and cost of manufacture. The experimental work described above immediately indicates the following possibilities, which should be borne in mind when considering any particular application :—

(1) Time-lag may be reduced by increasing the separating force, but it has been shown that to effect a substantial decrease in lag this force must be very large. If it is applied permanently it must not be greater than the initial attractive force when the current is switched on, and this is liable to be small owing to the presence of the air-gap when the clutch is open.

Fig. 5.



(2) Another method of reducing time-lag is to interpose a small gap between the plates by attaching a thin layer of friction material to one of them. The disadvantage of this method is the increase in magnetization current necessary to produce the same attractive force between the plates. However, the reduction in flux for the normal exciting

current was only 7 and 15 per cent. for the 0·001-inch and 0·005-inch air-gaps respectively, so, if layers of friction material of this order of thickness could be used, the decrease in time-lag might justify a slight decrease in maximum flux or alternatively a larger magnetizing current.

(3) The resistance of the eddy-current paths would be increased by lamination of the steel.

(4) The decay of flux could be accelerated by the injection of a reverse current into the coil as the main current is broken.

Of these methods the third is probably the most satisfactory, but would have a high initial cost.

Acknowledgements.

This work was carried out in the Electrotechnics Department of Manchester University, and the author gratefully acknowledges the helpful advice and facilities provided by Professor Willis Jackson.

Thanks are also due to Messrs. Leyland Motors, Ltd., who kindly loaned apparatus on which to conduct the experiments.

XXXVI. *Some Disputed Questions in the Philosophy of the Physical Sciences.*

By E. T. Whittaker *.

(From the Address of the President at the Annual Statutory Meeting of the Royal Society of Edinburgh, October 27, 1941.)

[Received March 2, 1942.]

More than two thousand years ago the Greek philosophers raised certain questions, which are still undecided, about the origin and character of knowledge regarding the external world. After a period of comparative quiet, the discussion has become very active recently, under the stimulus of the new discoveries in mathematical physics; and, in particular, a lively debate is in progress at the present moment between Sir Arthur Eddington and Dr. Harold Jeffreys of Cambridge, Professor Milne of Oxford, Sir James Jeans, and Professor Dingle of the Imperial College, the subject being the respective shares of reason and observation in the discovery of the laws of nature. I propose this afternoon to offer some remarks on the history and present state of this controversy.

It is admitted by everyone that the mathematical pre-calculation of natural events is conceivable only because the world is rationally made: and that the ideal of science is to discover laws and equations sufficient to make every physical event predictable, so that mankind will some day possess a corpus of purely mathematical relations, capable of representing and foretelling every happening in the inanimate external world. Now it is characteristic of any branch of mathematics, that the whole of it can be worked out from a few definitions and assumptions set down at the beginning. The question therefore arises, what are the fundamental data or postulates from which the complete set of laws of the material universe can be deduced by pure mathematics? And are these data or postulates furnished to us by the senses—by observation and experiment—or are they self-evident truths, revealed and assured to us by intuition?

This fundamental problem of the philosophy of nature was first conceived and discussed by the ancient Greeks, whose judgment on it can be very clearly seen in their treatment of geometry. Geometry is the science of spatial relations in the external world and so is essentially a part of natural philosophy: and the knowledge of many geometrical properties was, doubtless, originally derived from observation: thus, according to one conjecture, Thales became convinced that the angle in a semicircle is always a right angle, by gazing at tiles which showed, as part of the ornamental design, a rectangle inscribed in a circle. This earliest stage of geometry corresponded to that now generally imposed on school children, who are taught to draw figures with ruler and compasses, and to measure angles of triangles with protractors, before being introduced to the deductive treatment of the subject. When the logical connexion of different theorems was established by Pythagoras and his school, the idea gained ground that it might be possible to link up the entire science of geometry into a chain of propositions obtained by syllogistic reasoning from a small number of original premisses. The great question was, from whence were these premisses derived? Plato believed that they could all be obtained by pure intellection; and Euclid drew up a list of five "common notions" or axioms and five "postulates," from which he professed to demonstrate all the results of geometry as logical conclusions. The "common notions," such as "Things that are equal to the same thing are equal to each other," were presented as self-evident truths which belong alike to every science. The postulates, such as "Let it be granted that a circle may be described with any centre and diameter," were put forward as admitted feasibilities, peculiar to the science of geometry. Neither for common notions nor for postulates was any proof offered: the disciple was expected to know by intuition that they were necessary, that things could not be otherwise. Thus the Greek philosophers taught that although geometry was a science relating to the sensible external universe, it could be built up completely without having recourse at any stage to quantitative observation.

---

* Communicated by the Author.

Even in the ancient world there was some uneasiness about the validity of this doctrine. One of Euclid's premisses was the famous "parallel-postulate," "If a straight line, falling on two straight lines, makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than two right angles." This is not a particularly simple statement—are we justified in accepting it without proof as true? Some of the Greeks, and many later mathematicians down to the nineteenth century, were inclined to suspect that, though true, it was not a primary truth which could be referred immediately to the intuition, but a theorem which could be established by deductive reasoning from simpler self-evident assumptions. Accordingly they tried to discover some more plausible axiom from which it could be derived: an endeavour which met with a certain degree of success when in the seventeenth century Wallis showed that if the existence of triangles different in size but similar in shape could be assumed, then the parallel-postulate could be proved and Euclidean geometry re-constituted. Even this, however, was not altogether satisfactory: for it cannot be claimed that a belief in similar triangles is a necessity of human thought. In fact, there are no similar triangles in spherical geometry—that is to say, the mind can conceive of a type of geometry in which similar triangles do not exist. Thus Wallis's postulate is not something which we are compelled to believe by reason of the structure of our minds.

We are therefore forced to the conclusion that Euclidean geometry cannot be deduced from self-evident axioms; and hence the question as to whether it is the true geometry of the actual world can only be settled by making measurements and examining how far they agree with the theoretical predictions. Geometry is a branch of experimental knowledge. This fact, so important from the point of view of the philosophy of the physical sciences, obtained general acceptance very slowly—the belief that the Euclidean system is necessarily true was unquestioned for two millenniums, and became finally discredited only in the second half of the nineteenth century.

We have now to inquire whether the ancient Greeks thought it possible that the other sciences might, like geometry, be deduced logically from premisses self-evident to the intuition. On this point Aristotle, at any rate, was quite clear. "The principles which lie at the basis of any particular science," he says*, "are derived from experience (*ἐμπειρία*): thus it is from astronomical observation that we derive the principles of astronomical science." He poured scorn on what he called *ἀπειρία*†—that is, the state of those who devote themselves entirely to abstract reasoning from intuitive postulates, and are indifferent to facts. And he insisted on the vital importance of investigating any subject scientifically (*φυσικῶς*)—that is to say, by the study of sensible objects in nature‡—rather than dialectically (*λογικῶς*), that is, by pure deduction from unproved assumptions. The primary elements of true science must be the result of induction (*ἐπαγωγή*) based on phenomena, proceeding from particulars to universals. Sense-perception is indispensable to induction, and induction to discovery of the laws of nature.

The opinions of Aristotle on this subject were accepted by the schoolmen who developed his philosophy in the Middle Ages. St. Thomas Aquinas himself said § that the only secure foundation is experiment,

---

[*] 'Prior Analytics,' i. 30.

[†] *De Gen. et Corr.* i. 2.

[‡] *Ibid.*

[§] *Cf.* St. Thomas, "Commentary on Aristotle's 'Physics'," Lib. viii. Cap. 1, Lect. 3, 4.

and many of the scholastics made accurate observations and rational deductions from them. It may seem at first sight surprising that under these circumstances the thirteenth and fourteenth centuries did not anticipate the brilliant scientific discoveries of the seventeenth and eighteenth. The chief reasons were, I think, these. In the first place, the outstanding example of Euclidean geometry, at that time the most fully developed branch of science, whose results were believed to have been obtained without any dependence on observation or experiment, encouraged among many of the schoolmen a tendency to neglect the principles so clearly set forth by Aristotle and St. Thomas, and to concentrate their interest on pure deduction; and this was not uncongenial to their general outlook, for, recognizing in the external physical order an expression of the Divine Reason, they inferred that the human reason was the proper organ for its investigation. In the second place, it is to be remembered that Aristotle was a biologist rather than a physicist; and the biologist, at least in the earlier stages of biology, classifies rather than measures. The emphasis in Aristotelianism was decidedly on classification; and the idea of mathematical precision in quantitative measurements, which is the fundamental principle of modern physics, scarcely entered into the conception of Induction which the Scholastics inherited from the Peripatetics. Again, the primitive biologist depends on observation (the study of those appearances which nature presents of her own accord) rather than on experiment (the artificial production of effects under conditions specially selected and controlled by the experimenter). Simple observation will carry the zoologist a long way: but it is insufficient for the purposes, say, of the chemist, who has to disentangle all the antecedents and concomitants of the reaction he has observed. Hence Aristotle and his followers were much more at home in the biological than in the physical domain.

The great archetype of modern scientific discovery was the Newtonian law of gravitation. How wonderful an achievement it was may be seen by comparing it with the concept of nature as a system of vortices proposed only forty years earlier by Descartes, in which there is no attempt to obtain concordance between the deductions of the theory and experimental data. With Newton, on the other hand, we find the complete technique of research as it is practised to-day—namely, first the accurate measurement of phenomena; secondly, the imagination of a physical hypothesis likely to account for them, then the working out by mathematics of the detailed consequences of the hypothesis; and lastly, the comparison of the results of these calculations with the observations. When we reflect that by this investigation the entire cosmos was for the first time shown to conform to prediction based on a mathematical law, and that it provided the model and pattern for all subsequent progress, we cannot but confirm the judgment engraved below the statue in Trinity chapel: *Newton, qui genus humanum ingenio superavit.*

From the time of Newton until quite recently the principle that science rests fundamentally on observation and experiment has been unchallenged. In order to understand the present revolt against it, let us first consider some striking cases in which a mathematician, working in his study without any contact with laboratories, has been able to predict the existence of wholly new and unexpected phenomena in the external world.

A typical example is Hamilton's discovery of conical refraction. If we mark a dot on a piece of paper, and look at it through a crystal of Iceland spar, we see in general not one dot but two: this is because the spar has the property called *double refraction.* In 1821 the French physicist Fresnel discovered the equation of the *wave-surface*, or locus at any instant of a disturbance generated at a particular point at some previous instant, in a doubly refracting crystal; but he did not study the geometrical features of the surface as a mathematician would do. This was precisely what Hamilton did. He found that Fresnel's surface had some remarkable singularities, such as sharp peaks like the vertex of a cone, at each of which it had an infinite number of tangent-planes; and from the existence of these mathematical peculiarities in the wave-surface he inferred the existence of a corresponding optical phenomenon of a most amazing kind, namely, that a ray of light within the crystal would, under certain circumstances, be divided on emergence into an infinite number of rays, constituting a conical surface, while a single ray in air, incident on the crystal, might give a cone of rays inside the crystal. It should therefore be possible to arrange an experiment so that a dot on a sheet of paper, when viewed through a biaxal crystal, should appear not as two dots, but as a complete circle. This prediction was immediately verified.

Since the beginning of the twentieth century many novel and remarkable effects in physics have been discovered by mathematicians. Einstein's predictions of the bending of light-rays by the sun's attraction, and of the red-shift of spectral lines emitted in a strong gravitational field, are two notable instances. These were deduced from his general theory of relativity, which was entirely mathematical—that is to say, it did not originate in, or depend on, any new experiments or observations. Yet another striking discovery of the same type was the recognition that ordinary hydrogen gas is a mixture of two different kinds of molecules. When it is remembered that for a century and a half hydrogen has been familiar to every schoolboy beginning science, and has been the subject of innumerable experiments by highly trained investigators, the announcement of its composite character came as a great surprise. The mathematical reasoning involved turned on a distinction which is made in algebra: if we have two algebraic quantities x and y, we can form from them certain expressions such as $x+y$ whose value is unaltered when x and y are interchanged with each other; these are said to be *symmetric* in x and y. We can also form expressions such as $x-y$ whose value is reversed in sign when x and y are interchanged; these are said to be *skew* in x and y. Now it was found, in 1927, that the mathematical equations which represent the conditions of existence of the hydrogen molecule possess two different solutions, of which one is symmetric and the other is skew. It followed that there must be two different kinds of hydrogen molecule, to which the names *para-hydrogen* and *ortho-hydrogen* were given. These two tautomers behave in exactly the same way as regards the formation of chemical compounds, which explains why the chemists have never distinguished or separated them; but the specific heat of para-hydrogen is greater at low temperatures than that of ortho-hydrogen; and their boiling-points and conductivities are also different. Hydrogen gas prepared by the usual chemical processes consists of one-quarter para-hydrogen and three-quarters ortho-hydrogen.

We have now to consider whether the doctrine that science must be based on observation needs any modification, in view of the fact that discoveries in physics of the most unexpected kind and of the greatest importance are frequently made by mathematicians who have never performed, or even seen, an experiment in their lives. Obviously the dominant principle to be taken into account here is that the world is *rational*, the different phenomena being interconnected logically so that if we have found by observation a certain number of them, we can deduce the others by pure reasoning without making any fresh observations. Thus if we find by experiment that the coefficient of magnetization of a piece of soft iron decreases as its temperature increases, we can predict that a magnet will become heated when it is moved from weak to strong regions of the magnetic field.

We are thus led up to the most important problem of Natural Philosophy: What is the minimum set of observational data which is sufficient to form a basis for the whole edifice of physical theory? Certain eminent living physicists now give to this question the amazing answer: None at all! That is to say, they repudiate altogether the principle enunciated by Aristotle, St. Thomas, and Newton, and return to the attitude which was maintained for so long with regard to geometry, and which they now propose to extend to the whole of physical science.

In order to appreciate their standpoint, so unexpected and paradoxical as it must at first seem, let us consider one particular branch of physics, namely, the subject of electricity and magnetism. The early experiments and deductions in frictional electricity by Gilbert, Gray, Du Fay, Watson, Franklin, and Cavendish were completed by the measurement of the attraction between two electrically charged bodies, and its representation by the law generally known by the name of Coulomb, but actually first discovered by Priestley. In magnetism the recognition of magnetic poles by Petrus Peregrinus in the thirteenth century led eventually to the determination by Michell of the law of force between them. In the first half of the nineteenth century Oersted found that an electric current generates a magnetic field, Ampère showed that a ponderomotive force is exerted between two circuits carrying electric currents, and Faraday discovered the induction of currents. These researches, spread over six centuries and pursued incessantly during the last two of them, were experimental, and each of the physical properties concerned rested independently on its own observational basis. The various results were finally combined and consummated in the general electromagnetic theory of Maxwell, which comprehended all known electric and magnetic phenomena.

Before Maxwell, the story is one of performing experiments and devising formulæ to represent the results. But the post-Maxwellian period is wholly different in character. Maxwell's synthesis went beyond the experiments and predicted the existence of something that had never as yet been observed, namely, electromagnetic waves—what are now called wireless waves—and the correctness of this inference was verified after the death of Maxwell by the celebrated researches of Hertz. It may be said generally that experimental investigations in electricity from 1870 to 1900, such, for instance, as Rowland's work on fields produced by the motion of electric charges, were directly inspired by Maxwell's theory, and that the mathematical equations were capable of furnishing the results of the experiments beforehand.

The change in the method of discovery after Maxwell may be illustrated by a simple analogy. Suppose that a map of Scotland is pasted on stiff cardboard and then cut up into small irregular pieces, so that it can be used as a jigsaw puzzle. Anyone who tries to solve the puzzle does not at first know what is represented, and his only possibility of procedure is to find pieces which fit into each other and so constitute larger parts of the whole. After a time, however, he will have progressed sufficiently to be able to guess that what is represented is Scotland, and from that time onwards he completes the work not by finding pieces which fit into each other, but by using his *à priori* knowledge of Scotland to put every fragment into its proper place. These two methods may be likened to the two types of research in physical science: the earlier, proceeding step by step by experiment in special topics; and the later, knowing *à priori* what ought to be, because a guiding principle is now available for the whole, permitting the extension of knowledge by purely rational methods. When the work of fitting together the puzzle is completed, the picture alone remains as significant—that is to say, in physics, the mathematical theory.

The logical unity achieved in electrical science by Maxwell had this consequence, that the vast majority of the experiments, on which it had originally been built up, were now superfluous, and the number of ultimate independent observational facts necessary for its establishment was very small. Eventually they were reduced to only one, namely, that when a hollow metallic vessel is electrified there is no electric field in the air inside. From this single datum, combined with *à priori* principles such as the axiom of relativity, it is possible to derive mathematically, first, the inverse-square law of force between electric charges at rest, then (by considering these charges in motion relative to an observer) to deduce the existence of magnetic force, and finally to obtain in Maxwell's form the general equations of the electromagnetic field.

The single experimental fact on which electric and magnetic science can be founded may be put in the form, "It is impossible to set up an electric field in any region of space by enclosing the space in a hollow conductor of any shape or size and charging the outside of the conductor"; and in this form it bears a certain family likeness to other statements on which important branches of physics have been based, such as the postulate of thermodynamics (from which a great part of physical chemistry is derived), "It is impossible to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of the surrounding objects"; or the postulate of Relativity, "It is impossible to detect a uniform translatory motion, which is possessed by a system as a whole, by observations of phenomena taking place wholly within the system"; or the postulate (which plays an important part in the explanation of homopolar bonds in chemistry) that "It is impossible at any instant to assert that a particular electron is identical with some particular electron which had been observed at an earlier instant"; or the postulate of Imperfect Definition in quantum mechanics, "It is impossible to measure precisely the momentum of a particle at the same time as a precise measurement of its position is made." Each of these statements, which I propose to call *Postulates of Impotence*, asserts the impossibility of achieving something, even though there may be an infinite number of ways of trying to achieve it. A postulate of impotence is not the direct result of an experiment, or of any finite number of experiments; it does not mention any measurement, or any numerical relation or analytical equation; it is the assertion of a conviction of the mind, that all attempts to do a certain thing, however made, are bound to fail. We must therefore distinguish a postulate of impotence, on the one hand, from an experimental fact; and we must also distinguish it, on the other hand, from the statements of Pure Mathematics, which do not depend in any way on experience, but are necessitated by the structure of the human mind ; such a statement, for instance, as "It is impossible to find any power of two which is divisible by three." We cannot conceive any universe in which this statement would be untrue, whereas we can quite readily imagine a universe in which any physical postulate of impotence would be untrue.

It seems possible that while physics must continue to progress by building on experiments, any branch of it which is in a highly developed state may be exhibited as a set of logical deductions from postulates of impotence, as has already happened to thermodynamics. We may therefore conjecturally look forward to a time in the future when a treatise on any branch of physics could, if so desired, be written in the same style as Euclid's *Elements of Geometry*, beginning with some *à priori* principles, namely, postulates of impotence, and then deriving everything else from them by syllogistic reasoning.

With our minds thus prepared, let us approach the doctrines that Sir Arthur Eddington and Professor E. A. Milne have lately given to the world.

Milne begins by regarding the universe as an aggregate, of which one member is our galactic system of stars, the other members being the extra-galactic nebulæ, each of which is actually a vast system of stars. He represents this aggregate in an idealized fashion as a collection of particles, each particle corresponding to one nebula ; and each particle is supposed to carry an observer, equipped with a clock and a theodolite and with apparatus for sending and receiving light-signals. Milne then postulates that this ideal universe satisfies what he calls the *cosmological principle*, namely, that the map of the world which any particle-observer constructs from his own observations is identical with the map which any other particle-observer constructs from *his* own observations—that is, each sees the same sequence of views of the life-history of the universe.

The cosmological principle evidently bears a fairly close resemblance to the axiom of relativity, and may, like it, be expressed in the form of a postulate of impotence*. Thus what Milne does essentially is to assume a postulate of impotence and then deduce its logical consequences, much as the science of thermodynamics is deduced from the Second Law ; or, to go further back, much in the same way as Euclid deduced the whole of geometry from his common notions and postulates. Indeed, Milne describes his work as an attempt to do for systems of particles in motion what the Greeks did for static assemblages of points. He arrives at a detailed account of the life-history of the universe, and derives the laws of dynamics, the inverse-square law of gravitation, and the equations of electromagnetism.

---

* "It is impossible to tell where one is in the universe."

The work of Sir Arthur Eddington is, on the philosophical side, more extensive and much more difficult to characterize than that of Professor Milne. In two books, published in 1936 and 1939 respectively, he has set forth systematically what is offered as a complete philosophy of physical science.

The origin of his distinctive ideas may, I think, be traced to the circumstance that in the upbuilding both of relativity and of quantum mechanics the question, " What is it that we really observe ? " has played a great part. The critical examination arising out of this question showed that the older physicists had been accustomed to talk about things that are inaccessible to observation, whose existence and meaning are therefore highly doubtful. It came to be recognized, for instance, that no significance can be attached to the term " absolute velocity in space," and that it is impossible to imagine an experiment which could follow continuously the motion of an electron in an atom. These advances towards precision of thought, and the remarkable consequences which followed from them, made on Eddington a profound impression, and his aim in recent years has been to extend them, by making a thoroughgoing study of the nature and limitations of our knowledge of the physical sciences.

His attitude may be indicated by an analogy. Suppose a sea-fisherman invariably finds on examining his catch that no fish in it is less than two inches long. He might be inclined to assert as a fundamental law of zoology that all fishes are at least two inches in length, did he not know the real reason for the observed fact, namely, that he was fishing with a net of two-inch mesh. In interpreting the analogy, we may compare the net to the methods of research employed in physics, and the two-inch minimum for fishes to the laws physicists have discovered, which have in the past been regarded as describing properties of the real world, but which may be merely the necessary consequences of the procedure used in obtaining them. Hence Eddington's insistence on the importance of a proper investigation of the sensory equipment with which we observe, and the mental equipment by aid of which we formulate the results of observation. From this scrutiny he derives certain conclusions which he calls *epistemological principles*—principles relating to the method or ground of knowledge—and which, he asserts, are adequate to supersede entirely all observation and experiment as the basis of physical science.

This claim is so astounding that it must be stated in his own words. " I believe," he says *, " that the whole system of fundamental hypotheses [of physics] can be replaced by epistemological principles. Or to put it equivalently, all the laws of nature that are usually classed as fundamental can be foreseen wholly from epistemological considerations." "This means," he adds *, "that the fundamental laws and constants of physics are wholly subjective, being the mark of the observer's sensory and intellectual equipment on the knowledge obtained through such equipment, for we could not have this kind of *à priori* knowledge of laws governing an objective universe." "An intelligence," he says in another place †, "unacquainted with our universe, but acquainted with the system of thought by which the human mind interprets to itself the content of its sensory experience, should be able to attain all the knowledge of physics that we have attained by experiment. He would not deduce the particular events and objects of our experience, but he would deduce the generalizations we have based on them. For example, he would infer the existence and properties of sodium, but not the dimensions of the earth."

---

* 'The Philosophy of Physical Science,' 1939, p. 56.

The doctrine thus proclaimed is so entirely contrary to all the received ideas of the nature of science, that it will certainly not be accepted without a careful examination. In the first place, we may remark that Eddington's epistemological principles cannot be absolutely antecedent to all experience of the external world, for without some such experience it would be impossible to attach any meaning to the language employed (this indeed he admits). In the second place, we may observe that the notion of an "intuitive truth" has a sense that is relative rather than absolute. Take, for example, the statement that the sum of 1 and 5 is equal to the sum of 2 and 4. We are accustomed to speak of this as a proposition necessitated by pure logic, and indeed, to those who have mastered Whitehead and Russell's 'Principia Mathematica,' so it is; but it is not an intuitive truth to a savage who cannot count beyond four, and there is no doubt that it came to be accepted only as the result of a long development of culture, in which the operation was frequently performed of adjoining one object to five objects and observing that the same collection could be obtained by adjoining two objects to four objects. That is to say, a statement which we believe from inescapable necessity was originally a conviction arrived at from countless experiments. We must therefore envisage the possibility that the postulates of impotence and the Eddingtonian principles, to which most physicists, if they would assent to them at all, would assent on the ground of observation, may, by a more highly cultured or more penetrating mind, be perceived as self-luminous and irresistible truths, for which all proof

---

[*] *Ibid.* p. 104.

[†] 'Relativity Theory of Protons and Electrons,' 1936, p. 327.

would be superfluous. In Indian philosophy, I understand, there is a doctrine which asserts the possibility of an ineffable union with the infinite intelligence; but one hesitates to put forward, as the solution of the difficulties attending the epistemological assumptions, the suggestion that some such condition is already enjoyed by their distinguished author.

In order to examine more closely the validity of the thesis before us, let us consider in detail one of its principles, and the physical results which Eddington claims to have deduced from it. An assumption which he specially mentions * is the "special relativity principle." This, and some of his other epistemological axioms, happen to be closely connected with some of what I called "postulates of impotence"; but the concept of a postulate of impotence is quite different from Eddington's concept of an epistemological principle, since the postulates of impotence are generalizations from experiment, which epistemological principles expressly are not; and also because some of the latter are not assertions of inability to do anything.

Take, then, the special relativity principle, and a physical result which Eddington offers as a deduction from it. "As an example," he says †, "we may take the law of increase of mass with velocity, which has been the subject of many famous experiments. It is now realized that this law automatically results from the engrained form of thought which separates the fourfold order of events into a threefold order of space and an order of time. When knowledge is formulated in a frame which compels us to separate a time dimension from the fourfold order to which it belongs, a component called the mass is correspondingly separated from the fourfold vector to which it belongs: and it requires no very profound study of the conditions of separation to see how the separated component is related to the rest of the vector which prescribes the velocity. It is this relation which is rediscovered when we determine experimentally the change of mass with velocity."

This account of the matter ignores everything in the deduction of the mass-velocity formula except one particular step, namely, the separation of the temporal component from a vector in space-time. But, before this separation can be effected, we must know all about the vector—that is, we must already be in possession of relativist dynamics. Now relativist dynamics is based in part on the Axiom of Relativity, which may perhaps be called an epistemological principle (though actually it was arrived at only as an inference from an enormous number of experiments); but the

---

* 'The Philosophy of Physical Science,' p. 39.

† *Ibid.* p. 116.

Axiom of Relativity taken alone is certainly not sufficient to yield all that is wanted. There is not, so far as I know, any treatment of relativity-theory which does not in some way, directly or indirectly, make use of the assumption that the formulæ of relativist dynamics reduce to the formulæ of Newtonian dynamics when the velocities concerned are small. Indeed, without this assumption it does not seem possible to set up the desired connexion between the mathematical equations, on the one hand, and phenomena in the external world, on the other. But Newtonian dynamics was undoubtedly founded on experiment, and it is difficult to see that Eddington has provided any other derivation.

The perplexity generated by his account of the laws of nature is not lessened by studying his work on the fundamental constants. "Not only the laws of nature," he says*, "but the constants of nature can be deduced from epistemological considerations, so that we can have *à priori* knowledge of them." He actually obtains four constants, namely, the number of particles in the universe, the ratio of the electric to the gravitational force between a proton and an electron, the ratio of the mass of the proton to the mass of the electron, and what is called the fine-structure constant, which is of great importance in atomic physics. The last three of these were already known experimentally, but the first, the number of particles in the universe, which he calls the *cosmical number*, is hardly susceptible of observational verification; it is derived by epistemological reasoning of the purest kind, being in fact equated to the number of independent quadruple wave-functions, which is $2 \times 136 \times 2^{256}$. Naturally the question arises as to what exactly is meant by the "number of particles in the universe." At the time when the theory was originally propounded it was believed that all atoms are composed of two kinds of elementary particles, namely, protons and electrons; and Eddington defined the cosmical number to be simply the number of protons plus the number of electrons in the world. Since then the situation has become more complicated: a fresh elementary particle, the neutron, was found experimentally in 1932; another, the positron, theoretically in 1930 and experimentally in 1933; another, the neutrino, theoretically in 1931; and yet another, the meson, theoretically in 1935 and observationally in 1937. In the light of these discoveries, some re-interpretation of the cosmical number was evidently necessary. Eddington has decided† that a neutron should count as two particles, and a positron as minus one, so that the creation and annihilation of electrons and positrons in pairs will not affect the total. What is to be done about mesons and neutrinos is not yet settled. It must be said that readjustments of this kind are too suggestive of patching-up to be altogether satisfying. However, their author's confidence has in no way diminished, and he now claims * that the cosmical number determines the speed of recession of the distant nebulæ and the range of action of the forces between the particles in an atomic nucleus ; and that all these things, being accounted for epistemologically, are *ipso facto* subjective—they are demolished as part of the objective world †.

---

* 'The Philosophy of Physical Science,' p. 58.

† 'The Philosophy of Physical Science,' p. 170.

From the Greek nature-philosophers to Eddington the wheel has come full circle. Each of them believed, for example, that the value of the sum of the angles of a triangle is a necessary consequence of the constitution of the human mind. But they do not agree on what the value is. The Greeks said it was two right angles, while Eddington says that for triangles of astronomical size it is greater than two right angles by an amount depending on the size of the triangle.

A word may be said in conclusion on the question as to whether Eddington's general attitude may be subsumed under any of the historical philosophical schools. The two philosophers to whom he acknowledges himself indebted are Kant ‡ and Bertrand Russell §—an oddly-assorted pair, since Russell has devoted much of his activity to confuting Kant. What Eddington admires in the philosopher of Königsberg is the *à priori* doctrine regarding space, to which his own principles bear a certain resemblance. This affiliation is not likely to commend the new epistemology to mathematicians, who still remember with bitterness the harm that was done in the early days of non-Euclidean geometry by Kantian opposition ; and, perhaps partly on this account, Eddingtonian views have not hitherto made many converts. The admiration which is universally accorded to the genius of their author is assuredly not diminished by his stupendous attempt to base the study of the external world on a new foundation : but on the work itself the verdict as yet is *not proven.*

---

* *Ibid.* p. 177.
† *Ibid.* p. 59.
‡ *Ibid.* p. 188.
§ *Ibid.* p. 151.

XXXVII. *A Note on Prof. W. Wilson's " Dimensions of Physical Quantities."*

By G. Burniston Brown *.

[Received March 9, 1942.]

My friend Prof. Wilson has made a sharp raid on my position and bombs of varying calibre have been distributed. I cannot help thinking, however, that the damage is slight and that, in any case, I can take it.

In the first place, I should not have called my paper " A New Treatment of the Theory of Dimensions " had it not been really new, and consequently I remain unperturbed by references to custom that " has the sanction of Maxwell." Still less am I perturbed by attempts to mix my theory with that of Maxwell, as Prof. Wilson does with his formula $T=2\pi\sqrt{\lambda C}$: it is only to be expected that the result would be " surely nonsense." In the E.M. system, in which the formula holds, λ has, on my theory, the dimensions L, and C the dimensions $L^{-1}T^2$: the square root of the product gives T as it should. My definition of dimensions does *not* require the dimensions of a physical quantity to be the same on the E.S. and E.M. systems, since they symbolize different operations of measurement : that is one point of fundamental difference from Maxwell.

As regards the measurement of length, I have dealt fairly fully with this on page 423 †, but Prof. Wilson does not refer to it. In spite of the fact that the whole subject is usually entitled in the textbooks " Units and Dimensions," few physicists seem to realize that *the procedure that defines the unit defines the dimensions*. The comparison of two lengths and the measurement of a length are quite different processes and the results are expressed differently. When we *compare* two lengths we perform *two* operations of measurement (Prof. Wilson's 394 inches and 39·4 inches), and when we form the ratio the dimensions cancel just as the units do, and the result is correctly expressed as " 10 " (not " 10 metres " as Prof. Wilson seems to assume). One length is ten times the other, whatever the units. When, however, we *measure* a length we perform only *one* operation of measurement, since we cannot measure the standard (except in terms of another standard, which is only evading the issue). The result of this procedure is correctly expressed as "10 metres" and dimensionally as 10 L; where there are units there are dimensions.

---

* Communicated by the Author.

† Proc. Phys. Soc. liii. p. 418 (1941).

K and μ are dimensionless, just as they are unitless, for the simple reason that they are defined as *ratios*. K is not defined from Prof. Wilson's equation (6); if it were, we should not only never find the dimensions of electric charge, we should never find the unit.

With reference to the table of equivalent dimensions given at the end of my paper, Prof. J. B. S. Haldane has pointed out to me that angular momentum, though subject to conservation, does not fit into the table. I had hoped that some significance might have been attached to this table, but this would seem to make it unlikely.

My theory of dimensions is the result of adopting a philosophy of science quite different from that of Maxwell and classical physicists generally. By examining carefully how numbers are put into nature by the physicist, I hope to account for some of the numbers that come out. In this way I have given an explanation of the continual curious appearance of the velocity of light in situations where nothing optical is observed or measured. It is hardly necessary to point out that. in order to account in this way for well-known numbers, it is essential to consider what the physicist *does*, not what he *might do*. That is why, with definite units, we must have definite dimensions, or, to use Prof. Wilson's words, "unique" dimensions

XXXVIII. *The Theory of the Electric Charge and the Quantum Theory.—* Part III. †

By H. T. Flint, Reader in Physics,
University of London, King's College ‡.

[Received February 4, 1942.]

In the attempt to express the notation of the quantum theory in relativistic form it has been found that a five-dimensional system of co-ordinates provides a convenient mode of description; it is indeed forced upon us by the requirements of the quantum theory. The results obtained satisfy all the requirements of invariance in the sense of the theory of general relativity with a restricted use of the fifth co-ordinate. The character of this restriction is controlled by the application to the quantum theory. A constant, α, described as Kaluza's constant, occurs in the identification of some of the coefficients of the line element

$$d\sigma^2=\gamma_{\mu\nu}dx^\mu dx^\nu$$

with the electromagnetic scalar and vector potential.

This identification occurs in the following way:

$$\gamma_{m5}=\alpha\phi_m,$$

where $m=(1, 2, 3 \text{ or } 4)$.

This constant, together with c, h, and e with their usual meanings are the constants of the theory.

Quantization of the electric charge is incorporated in the theory and the concept of the mass of the charged particle, which is equal to $e/\alpha c^2$, is introduced because the equation of the geodesic of the continuum takes the form

$$\text{force}=\text{mass}\times\text{acceleration}$$

with rest mass $m_0=e/\alpha c^2$. The fundamental lengths h/m_0c and e/m_0c^2 occupy a natural place in the theory.

It has been found possible to develop field equations(1) for the electric charge on the principle that they are of the Maxwell type and by assuming that the covariant field tensor is expressible in the form of the curl of a vector. This treatment is, however, classical in character, the process of quantization in the modern sense not being introduced. To this point the theory is satisfactory as a starting-point for quantization because the concept of quantization of charge and the fundamental lengths already occupy a place in the classical background.

---

† For Parts I. and II. see Phil. Mag. (7) xxix. pp. 330 and 417 (1940).

‡ Communicated by the Author.

It has been shown that the process of quantization is to be undertaken by an appeal to geometric form, but with the introduction of new elements into geometry, the chief of which is the concept of matrix length.

The expression of the field equations in this framework involves the use of a notation suited to it, in particular the operation of the differentiation of matrices and the extension of the concept of such operators as div. and curl requires consideration.

The Development of the Notation.

In taking the first step in this direction it appears that the fundamental equation of the quantum theory, that is to say, Dirac's equation for the electron in an electromagnetic field, generalized to include a gravitational field, can be regarded as the gauging equation of matrix geometry. In the course of the development of this point of view the quantum matrices (γ^{μ}) are shown to satisfy the relation

$$\frac{\partial \gamma^{\mu}}{\partial x^{\nu}}+\gamma^{\rho}\Delta^{\mu}_{\rho\nu}=\kappa_{\nu}^{\mu}. \qquad (1)$$

In this relation any suffix occurring twice, as ρ in the second term of equation (1), is summed over all its possible values. In vector and tensor quantities Greek letters are used for suffixes with the values (1–5) and Latin suffixes for those with values (1–4). A vector, tensor or matrix quantity is represented by writing down a typical component. Thus (γ^{μ}) denotes all the fundamental matrices, i. e., $\mu=1, 2, 3, 4, 5$.

It will be necessary to distinguish between tensor components and matrix components, and in general Greek letters will be used to denote the type of variance of a tensor, and a special note must be added to avoid confusion where it may occur.

The quantity $\Delta^{\mu}_{\rho\nu}$ is a component of the five-dimensional Christoffel bracket expressions.

On account of the relations satisfied by the matrices, viz. :—

$$\gamma^{\mu}\gamma^{\nu}+\gamma^{\nu}\gamma^{\mu}=2\gamma^{\mu\nu}, \qquad (2)$$

it follows that

$$\kappa_{\nu}^{\mu}=\Delta_{\nu}\gamma^{\mu}-\gamma^{\mu}\Delta_{\nu}. \qquad (3)$$

Equation (1) now becomes

$$\frac{\partial \gamma^{\mu}}{\partial x^{\nu}}=-\gamma^{\rho}\Delta^{\mu}_{\rho\nu}+\Delta_{\nu}\gamma^{\mu}-\gamma^{\mu}\Delta_{\nu}. \qquad (4)$$

This gives the change in the component γ^{μ} corresponding to a parallel displacement (dx^{ν}). The first term on the right occurs in accordance with the Riemannian system of geometry, for the matrix γ^μ has the property of a contravariant vector. The additional terms occur because the quantity is also a matrix operator, and they suggest the place of the quantities (Δ_ν) in the theory. The place of (Δ_ν) in this calculus can be readily described by expressing a matrix in the form of a linear vector operator. In familiar vector analysis a linear vector operator A is expressed in the form

$$A = A_{mn} i_m i_n, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where the unit vectors of reference are (i_m).

For the present purpose it is necessary to introduce a set of unit vectors (i_m) and a companion set (ϵ_m) known as conjugates to the first set. Any vector, R, is represented in two ways, either by

$$R = R_m i_m$$

or by means of its conjugate

$$R^* = R^*_m \epsilon_m,$$

where $R_m{}^*$ is the conjugate complex quantity of R_m. Scalar operations such as $(i_m . i_n)$ are not contemplated but are replaced by

$$i_m . \epsilon_n = \delta_{mn}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The scalar product of vectors R and S is made by combining R and S* or R* and S. But

$$RS^* \neq R^*S.$$

In the same way the linear vector operator is now represented by

$$A = A_{mn} i_m \epsilon_n. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

It will be supposed that a set of vectors (i_m), (ϵ_n) belongs to each point of space and that vectors and operators are represented by reference to them.

It is necessary to distinguish between the vector component in the expression $R_n i_n$ and the component of a covariant vector. In order to do this we shall describe the vector R as a vector in the operator field. The matrices (Δ_ν) serve the purpose of relating operator vectors at neighbouring points. An operator field vector at a point $P(x^\nu)$ is represented by $R_n i_n$ and is regarded as equal to the vector $(R_n + dR_n)$ at the neighbouring point Q $(x^\nu + dx^\nu)$, provided that

$$dR_n = (\Delta_\nu)_{nm} R_m dx^\nu. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

This can be described as the rule for parallel displacement in the operator field.

If the appropriate vectors at Q are $(i_n + di_n)$ it is implied that

$$R_n i_n = (R_n + dR_n)(i_n + di_n), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

provided that dR_n is given by equation (8).

If this method of correlation at the two points is to be of general validity, it follows that for all vectors (R_n)

$$R_n \, di_n + dR_n i_n = 0. \qquad (10)$$

Whence

$$di_n = -(\Delta_\nu)_{mn} i_m \, dx^\nu. \qquad (11)$$

Since

$$\Delta_\nu = -(\Delta_\nu)_{mn} i_m \epsilon_n,$$

the right-hand side of this equation may be written

$$di_n = -\Delta_\nu . i_n \, dx^\nu. \qquad (12)$$

Changes of these magnitudes are thus required by the different standards of reference at the different points and are in this sense not to be regarded as actual changes.

If the vector R is a function of the co-ordinates, a component R_n on reaching Q becomes $\dfrac{\partial R_n}{\partial x^\nu} dx^\nu$ and the net change as reckoned at Q is

$$\left\{\frac{\partial R_n}{\partial x^\nu} - (\Delta_\nu)_{nm} R_m\right\} dx^\nu. \qquad (13)$$

A similar argument applies to the conjugate vector, but since it is desired to preserve the invariance of RS^* in a parallel displacement, the change in this case is

$$dS^*_n = -(\Delta_\nu)_{mn} S^*_m \, dx^\nu \qquad (14)$$

and

$$d\epsilon_n = (\Delta_\nu)_{nm} \epsilon_m \, dx^\nu = \epsilon_n \Delta_\nu \, dx^\nu. \qquad (15)$$

The Value of the Operators (Δ_ν).

In seeking the value of these operators we can refer to equation (4). In the form of an operation on the covariant quantities (γ_μ) this becomes

$$\frac{\partial \gamma_\mu}{\partial x_\nu} = \gamma_\rho \Delta^\rho_{\mu\nu} + \Delta_\nu \gamma_\mu - \gamma_\mu \Delta_\nu. \qquad (16)$$

The case of importance is that when an electromagnetic field exists in the absence of any appreciable gravitational field. In this case the following relations hold: (II. eqns. 1–4)

$$\gamma_5 = \alpha_0, \quad \gamma^m = \alpha^m, \quad \gamma_m = \alpha_m + \alpha\phi_m\alpha_0, \quad \gamma^5 = \alpha_0 - \alpha\phi_m\alpha^m. \qquad (17)$$

Since there is no gravitational field, $\alpha^m = \alpha_m$, and these matrices together with α_0 are those of Dirac satisfying the relations

$$\alpha_m\alpha_n + \alpha_n\alpha_m = 2\delta_{mn}, \quad \alpha_0\alpha_m + \alpha_m\alpha_0 = 0, \quad \alpha_0{}^2 = 1. \qquad (18)$$

In this case

$$\Delta^l{}_{mn} = \tfrac{1}{2}\alpha^2(\phi_m H_{ln} + \phi_n H_{ln}),$$

$$\Delta^5{}_{mn} = \tfrac{1}{2}\alpha\left(\frac{\partial \phi_n}{\partial x_m} + \frac{\partial \phi_m}{\partial x^n}\right) - \tfrac{1}{2}\alpha^3\phi_l(\phi_m H_{ln} + \phi_n H_{lm}).$$

$$\frac{\partial \gamma_m}{\partial x_n} = \alpha \frac{\partial \phi_m}{\partial x_n} \alpha_0,$$

where $$\mathrm{H}_{ln} = \frac{\partial \phi_l}{\partial x^n} - \frac{\partial \phi_n}{\partial x^l}.$$

It can now be verified that a solution of equation (16) is given by

$$\Delta_n = \tfrac{1}{4}\alpha(\alpha\phi_n \mathrm{H}_{lp}\alpha_p\alpha_l + \mathrm{H}_{ln}\alpha_0\alpha_l). \qquad \text{(19)}$$

This equation is in the form

$$\Delta_n = \mathrm{A}_n + \alpha\phi_n \mathrm{A}_0,$$

where $$\mathrm{A}_0 = \tfrac{1}{4}\alpha \mathrm{H}_{lp}\alpha_p\alpha_l, \quad \mathrm{A}_n = \tfrac{1}{4}\alpha \mathrm{H}_{ln}\alpha_0\alpha_l.$$

From the equations which relate four and five-dimensional quantities, it appears that the four-dimensional operators Γ_n and Γ_0 are given by

$$\Delta_5 = \Gamma_0 = \tfrac{1}{4}\alpha \mathrm{H}_{lp}\alpha_p\alpha_l, \quad \Gamma_n = \tfrac{1}{4}\alpha \mathrm{H}_{ln}\alpha_0\alpha_l. \qquad \text{(20)}$$

The form given to Δ_5 by this relation can be shown to satisfy equation (16) for the case $\nu=5$, when the left-hand side vanishes.

It should also be noted that in deriving the relation (19) we have left out a multiple of the unit vector which could be added to the right-hand side leaving equation (16) satisfied.

The operators occurring in this way are evidently associated with "spin," for on account of the asymmetry of H_{lp} in l and p, it is possible to replace $\mathrm{H}_{lp}\alpha_p\alpha_l$ by $\mathrm{H}_{lp}s_{pl}$, where $s_{pl} = \tfrac{1}{2}(\alpha_p\alpha_l - \alpha_l\alpha_p)$.

Thus $\Delta_5 = \tfrac{1}{4}\alpha \mathrm{H}_{lp}s_{pl}$ is a term corresponding to the expression for the energy of a doublet in an electromagnetic field and occurs, as is well known, in Dirac's theory of the electron. These parallel displacement operators may for this reason be described as spin operators.

It may be noted that the relation (16) now becomes an identity and corresponds to that which is satisfied by the components g_{mn} or $\gamma_{\mu\nu}$ in the case of covariant differentiation.

The Gauging Equation.

In the development of this equation a quantity ψ was introduced which was a scalar in the sense of Riemannian geometry but a vector in the sense of the operator field. A relation for this quantity was sought under the conditions that

$$\gamma^\mu \mathrm{C}_\mu \psi$$

remained unchanged when any ordinary vector (C_μ) was subjected to a parallel displacement.

The variation of ψ was denoted by $\frac{\partial \psi}{\partial x^\nu} dx^\nu$, and it was found that the requirement was fulfilled if the equation $\gamma^\mu \frac{\partial \psi}{\partial x^\mu} = 0$ is satisfied.

This equation is Dirac's equation provided that x^5 occurs in the way described.

The change in ψ in this procedure is due to a parallel displacement of this vector, consequently,

$$\frac{\partial \psi_n}{\partial x^\nu}=(\Delta_\nu)_{nm}\psi_m,$$

or in the vector form

$$\frac{\partial \psi}{\partial x^\nu}=\Delta_\nu\psi.$$

Thus the gauging equation can be expressed in the form

$$\gamma^\mu\Delta_\mu\psi=0. \quad \ldots \ldots \ldots \quad (21)$$

In order to see if this relation is compatible with the form obtained for (Δ_μ), we can again examine the case where the gravitational field is absent. In this case the relations (17) may be applied, and it follows that

$$\gamma^\mu\Delta_\mu=\alpha_m\Delta_m+(\alpha_0-\alpha\phi_n\alpha)_n\Delta_5=0.$$

The terms in this expression cancel in pairs on account of the properties of the Dirac matrices.

The Parallel Displacement of Operators.

An operator such as $A=A_{mn}i_m\epsilon_n$ can also be considered as subject to a parallel displacement of a magnitude which depends on the relations established in the case of the parallel displacement of vectors.

At two neighbouring points the above A and the displaced A, viz.

$$(A_{mn}+dA_{mn})(i_m+di_m)(\epsilon_n+d\epsilon_n),$$

are regarded as equal if the fundamental vectors change according to the relations (12) and (15).

It readily follows that this condition of equality requires that the parallel displacement dA be given by

$$\frac{\partial A}{\partial x^\nu}=\Delta_\nu A-A\Delta_\nu. \quad \ldots \ldots \ldots \quad (22)$$

If A is itself explicitly dependent on the co-ordinates we can, by the same argument as was employed for vectors, describe as the change of A,

$$\frac{dA}{dx^\nu}=\frac{\partial A}{\partial x^\nu}+A\Delta_\nu-\Delta_\nu A. \quad \ldots \ldots \ldots \quad (23)$$

It will be convenient to denote this variation by the symbol $D_\nu A$.

Vectors represented as Operators.

The matrices occurring in the theory are in general square matrices of 4 rows, but it is sometimes convenient to represent a matrix vector as a 5-rowed matrix in which only terms such as ψ_{n5} or ψ_{5n} appear. ($n=1, 2, 3, 4.$)

In this case the parallel displacement

$$d\psi=(\Delta_\nu\psi-\psi\Delta_\nu)\,dx^\nu$$

becomes

$$d\psi=\Delta_\nu\psi\,dx^\nu \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (24)$$

if $\psi=(\psi_{n5})$, since Δ_ν contains no component $(\Delta_\nu)_{5n}$, and consequently $\psi\Delta_\nu$ vanishes.

If

$$\psi=(\psi_{5n}),$$

$$d\psi=-\psi\Delta_\nu dx^\nu. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (25)$$

Thus the former changes like an ordinary field vector and the latter like the conjugate.

Thus (ψ_{5n}) and (ψ_{n5}) can be regarded as a conjugate pair.

A notation of the character described here has already been introduced and discussed by Schroedinger (2) in establishing Dirac's equation in a gravitational field. He applies his results to a four-dimensional continuum and shows that the similarity transformation requires the introduction of such operators as are introduced here.

This enables us to proceed with the notation in the knowledge that both this transformation and the requirements of relativistic invariance are satisfied.

Quantization of Field Equations.

By means of the notation just described an attempt will be made to set up field equations of the Maxwell type. The form of the equations will be that described in Part II., but the quantities occurring in the equations which correspond to the scalar and vector potentials and to the field components of Maxwell's equations become operators in the new equations. The process of differentiation will be modified to take account of the double character of the operators, *i. e.*, of the fact that they are vectors and tensors in the ordinary sense of these terms as well as operators.

The application of this procedure ensures the satisfaction both of relativistic invariance and of the similarity transformation and may be described as quantization of the field equations. The process described hitherto has been that of converting classical field equations into Hamiltonian form in order to recognize the appropriate mechanically conjugate quantities upon which to superpose the quantum conditions.

An example of this process is to be found in Yukawa's theory of the heavy electron. In this case a set of field equations identical with those referred to (3) formed the basis of the theory and were shown to be derivable from a Lagrangian function dependent on field quantities.

The Field Equations.

It will be assumed that the field can be described by means of a tensor $(T_{\mu\nu})$ subject to the condition which corresponds to the relation

$$T^{\mu\nu}{}_{;\nu}=0. \qquad (26)$$

of relativistic theory.

The symbol $;\nu$ means covariant differentiation, and summation over ν for the values (1, 2, 3, 4, 5) is implied.

In the case where T is a matrix quantity this becomes

$$T^{\mu\nu}{}_{;\nu}+T^{\mu\nu}\Delta_\nu-\Delta_\nu T^{\mu\nu}=0$$

or

$$\frac{1}{\sqrt{\gamma}}\frac{\partial}{\partial x^\nu}(\sqrt{\gamma}T^{\mu\nu})+T^{\mu\nu}\Delta_\nu-\Delta_\nu T^{\mu\nu}=0, \qquad (27)$$

γ denotes the determinant made up of the components $\gamma_{\mu\nu}$ and is placed equal to unity; moreover, $T^{\mu\nu}$ is antisymmetric in μ and ν. Thus the equation takes the form

$$\frac{\partial T^{\mu\nu}}{\partial x^\nu}+T^{\mu\nu}\Delta_\nu-\Delta_\nu T^{\mu\nu}=0. \qquad (28)$$

The form of $T_{\mu\nu}$ in the classical theory is

$$T_{\mu\nu}=\frac{\partial\psi_\nu}{\partial x^\mu}-\frac{\partial\psi_\mu}{\partial x^\nu}, \qquad (29)$$

but in the case of the use of operators this becomes

$$T_{\mu\nu}=\left(\frac{\partial\psi_\nu}{\partial x^\mu}-\psi_\nu\Delta_\mu-\Delta_\mu\psi_\nu\right)-\left(\frac{\partial\psi_\mu}{\partial x^\nu}+\psi_\mu\Delta_\nu-\Delta_\nu\psi_\mu\right). \qquad (30)$$

It will be convenient to describe this by means of the notation

$$T_{\mu\nu}=D_\mu\psi_\nu-D_\nu\psi_\mu. \qquad (31)$$

The correlation of the 5-dimensional components of T with their 4-dimensional counterparts is made by means of the equations quoted in Part II. (equations 5–10). The results may be compared with equations (20) of Part II., but there are some differences.

It is again assumed that (ψ_μ) has 4 components only and can consequently be regarded either as a four- or five-dimensional vector, in the latter case with $\psi_5=0$. Thus $T_{55}=0$, whence $A_{00}=0$.

$$T_{m5}=-\sigma\psi_m-\psi_m\Gamma_0+\Gamma_0\psi_m, \quad T_{5m}=-T_{m5}=A_{0m}=-A_{m0}$$

since $A_{00}=0$.

Thus

$$A_{0m}=\sigma\psi_m+\psi_m\Gamma_0-\Gamma_0\psi_m. \qquad (32)$$

For the contravariant components we have

$$T^{m5}=A^{m}{}_{0}-\alpha\phi_{n}A^{mn},$$

and since it is not necessary to consider the existence of a gravitational field this may be written

$$T^{m5}=A_{m0}-\alpha\phi_{n}A_{mn}.$$

By comparison with the field equations for the non-quantized case it appears that a complication will arise on account of the form of A_{m0}, which is no longer simply equal to $-\sigma\psi_m$ (equation (32)). In order to avoid this complication the condition

$$\psi_{m}\Gamma_{0}-\Gamma_{0}\psi_{m}=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

will be imposed. This makes $A_{m0}=-\sigma\psi_m$, and we thus keep closely to the form of the theory described in Part II.

Following the same methods the case $\mu=m$ is first considered. The appropriate field equation (*cf.* Part II. (19)) is

$$\frac{\partial T^{mn}}{\partial x^{n}}+\sigma T^{m5}+T^{mn}\Delta_{n}-\Delta_{n}T^{mn}+T^{m5}\Delta_{5}-\Delta_{5}T^{m5}=0.$$

On making the appropriate substitutions this becomes

$$\left(\frac{\partial}{\partial x^{n}}-\alpha\sigma\phi_{n}\right)A_{mn}+A_{mn}\Gamma_{n}-\Gamma_{n}A_{mn}=\sigma^{2}\psi_{m}. \quad . \quad . \quad . \quad (34)$$

In this way the operator $\left(\frac{\partial}{\partial x^{n}}-\alpha\sigma\phi_{n}\right)$ appears naturally in the analysis, and the frequent occurrence of the operation represented on the left-hand side of this equation makes it convenient to denote it by $\delta_n A_{mn}$. Thus the equation satisfied by a component of the field quantities (A_{mn}) which correspond to the field intensities of electromagnetic theory is

$$\delta_{n}A_{mn}=\sigma^{2}\psi_{m}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

This is a result which should be compared with equation (20) of Part II. It is in order to preserve this form that the relation (33) was introduced.

Proceeding now to the case $\mu=5$, we obtain

$$\frac{\partial T^{5n}}{\partial x^{n}}+T^{5n}\Delta_{n}-\Delta_{n}T^{5n}=0$$

since T^{55} vanishes.

Again, by making the appropriate substitutions this becomes

$$\frac{\partial A_{n0}}{\partial x^{n}}+\frac{1}{2}\alpha H_{nm}A_{nm}-\alpha\phi_{m}\frac{\partial A_{nm}}{\partial x^{n}}+(A_{n0}\Gamma_{n}-\Gamma_{n}A_{n0})-\alpha\phi_{m}(A_{nm}\Gamma_{n}-\Gamma_{n}A_{nm})=0.$$

By substitution for $\frac{\partial A_{nm}}{\partial x^n}$ from equation (34), and after substituting $A_{n0}=-\sigma\phi_n$, we obtain

$$\left(\frac{\partial}{\partial x^n}-\alpha\sigma\phi_n\right)\psi_n+\psi_n\Gamma_n-\Gamma_n\psi_n=\frac{1}{2}\frac{\alpha}{\sigma}H_{nm}A_{nm} \quad . \quad . \quad . \quad (36)$$

or

$$\delta_n\psi_n=\frac{1}{2}\frac{\alpha}{\sigma}H_{nm}A_{nm}\,, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

corresponding to equation (26) of Part II.

Similarly, for the conjugate quantities we obtain

$$\delta_n{}^*A^*{}_{mn}=\sigma^2\psi^*_m \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

and

$$\delta_n{}^*\psi_n{}^*=-\frac{1}{2}\frac{\alpha}{\sigma}H_{nm}A^*_{nm} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

where $\delta_n{}^*$ denotes that in the operator the positive sign is placed before the imaginary quantity σ.

A characteristic feature of the classical field theory is that

$$\frac{\partial T_{\mu\nu}}{\partial x^\lambda}+\frac{\partial T_{\nu\lambda}}{\partial x^\mu}+\frac{\partial T_{\lambda\mu}}{\partial x^\nu}=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

is an identity when the tensor $(T_{\mu\nu})$ is of the form (29). The form taken when the operator form (30) is adopted requires examination.

Each term of equation (40) takes the form

$$\frac{\partial T_{\mu\nu}}{\partial x^\lambda}+T_{\mu\nu}\Delta_\lambda-\Delta_\lambda T_{\mu\nu},$$

and the result of the relation (30) on combining the three terms represented in (40) is

$$(P_{\nu\mu}\psi_\lambda-\psi_\lambda P_{\nu\mu})+(P_{\lambda\nu}\psi_\mu-\psi_\mu P_{\lambda\nu})+(P_{\mu\lambda}\psi_\nu-\psi_\nu P_{\mu\lambda}) \quad . \quad . \quad . \quad (41)$$

where

$$P_{\nu\mu}=\frac{\partial\Delta_\mu}{\partial x^\nu}-\frac{\partial\Delta}{\partial x^\mu}+\Delta_\mu\Delta_\nu-\Delta_\nu\Delta_\mu. \quad . \quad . \quad . \quad . \quad . \quad (42)$$

The value of $P_{\nu\mu}$ can be derived from

$$\frac{\partial\gamma_\mu}{\partial x^\nu}=\gamma_\rho\Delta^\rho{}_{\mu\nu}+\Delta_\nu\gamma_\mu-\gamma_\mu\Delta_\nu, \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

which is the relation for the covariant (γ_μ) corresponding to equation (4) for the contravariant (γ^μ).

The equality $$\frac{\partial^2\gamma_\mu}{\partial x^\lambda\partial x^\nu}=\frac{\partial^2\gamma_\mu}{\partial x^\nu\partial x^\lambda}$$

leads to the result

$$\mathrm{B}^{\rho}_{\sigma\mu\lambda}\gamma_{\rho}=\mathrm{P}_{\lambda\mu}\gamma_{\sigma}-\gamma_{\sigma}\mathrm{P}_{\lambda\mu} \quad . \; . \; . \; . \; . \; . \; . \quad (44)$$

whence

$$\mathrm{P}_{\lambda\mu}=-\tfrac{1}{4}\mathrm{B}_{\rho\lambda\mu\sigma}s^{\rho\sigma}+f_{\lambda\mu}\mathrm{I}, \quad . \; . \; . \; . \; . \; . \; . \quad (45)$$

where $(\mathrm{B}_{\rho\lambda\mu\sigma})$ denotes the Riemannian Christoffel tensor for the continuum and $f_{\lambda\mu}$ is a multiple of the unit operator I. $s^{\rho\sigma}$ is the antisymmetric part of the product $\gamma^{\rho}\gamma^{\sigma}$

$$s^{\rho\sigma}=\tfrac{1}{2}(\gamma^{\rho}\gamma^{\sigma}-\gamma^{\sigma}\gamma^{\rho}). \quad . \; . \; . \; . \; . \; . \; . \quad (46)$$

Thus (41) takes the form

$$\tfrac{1}{4}\{\mathrm{B}_{\rho\mu\nu\sigma}(s^{\rho\sigma}\psi_{\lambda}-\psi_{\lambda}s^{\rho\sigma})+\mathrm{B}_{\rho\nu\lambda\sigma}(s^{\rho\sigma}\psi_{\mu}-\psi_{\mu}s^{\rho\sigma})+\mathrm{B}_{\rho\lambda\mu\sigma}(s^{\rho\sigma}\psi_{\nu}-\psi_{\nu}s^{\rho\sigma}\}. \quad (47)$$

This vanishes if $\mathrm{B}_{\rho\mu\nu\sigma}=0$, *i. e.*, if the space is Euclidean, otherwise some special relation is required between $(s^{\rho\sigma})$ and (ψ_{λ}). If, for example, these quantities commute, the expression vanishes.

An interesting point is that, if $\psi_{\lambda}=\gamma_{\lambda}$, the expression vanishes on account of the relations

$$\gamma_{\rho}s^{\mu\nu}-s^{\mu\nu}\gamma_{\rho}=2(\delta^{\mu}_{\rho}\gamma^{\nu}-\delta^{\nu}_{\rho}\gamma^{\mu}) \quad . \; . \; . \; . \; . \; . \quad (48)$$

and

$$\mathrm{B}_{\mu\nu\sigma\rho}+\mathrm{B}_{\mu\sigma\rho\nu}+\mathrm{B}_{\mu\rho\nu\sigma}=0. \quad . \; . \; . \; . \; . \; . \quad (49)$$

The Energy Tensor.

As in the previous case (Part II. p. 423) a tensor $\mathrm{P}^{\mu}{}_{\nu}$ is introduced, defined by

$$\mathrm{P}^{\mu}{}_{\nu}=\mathrm{T}^{*\mu\alpha}\mathrm{T}_{\nu\alpha}+\mathrm{T}^{\mu\alpha}\mathrm{T}^{*}{}_{\nu\alpha}. \quad . \; . \; . \; . \; . \; . \quad (50)$$

In this case T is defined by means of equation (30).

The field tensor, $\mathrm{R}^{\mu\nu}$, is formed, being defined by the relation

$$\mathrm{R}^{\mu\nu}=\mathrm{P}^{\mu\nu}-\tfrac{1}{4}\gamma^{\mu\nu}\mathrm{P}, \quad . \; . \; . \; . \; . \; . \; . \quad (51)$$

where

$$\mathrm{P}=\mathrm{P}^{\mu}{}_{\mu}.$$

The law of the field is expressed by

$$\mathrm{R}^{\mu\nu}{}_{;\,\nu}+\mathrm{R}^{\mu\nu}\Delta_{\nu}-\Delta_{\nu}\mathrm{R}^{\mu\nu}=0. \quad . \; . \; . \; . \; . \; . \quad (52)$$

If this be considered in a way analogous to that applied in the case of classical field equations (for example, as in Part II. p. 424), the first term becomes, for the case $\mu=m$, $(\mathrm{R}^{mn}{}_{;\,n}-\alpha\mathrm{H}^{m}{}_{n}\mathrm{R}_{5}{}^{n})$, where the Latin affixes imply that the operations are to be regarded as four-dimensional. The remaining terms $(\mathrm{R}^{m\nu}\Delta_{\nu}-\Delta_{\nu}\mathrm{R}^{m\nu})$ may be shown to be equivalent to

$$\mathrm{R}^{mn}\Gamma_{n}-\Gamma_{n}\mathrm{R}^{mn}+\Gamma_{0}\mathrm{R}_{5}{}^{m}-\mathrm{R}_{5}{}^{m}\Gamma_{0}.$$

In this case equation (52) becomes

$$(\mathrm{R}^{mn}{}_{n}+\mathrm{R}^{mn}\Gamma_{n}-\Gamma_{n}\mathrm{R}^{mn})=\alpha\mathrm{H}^{m}{}_{n}\mathrm{R}_{5}{}^{n}+\mathrm{R}_{5}{}^{m}\Gamma_{0}-\Gamma_{0}\mathrm{R}_{5}{}^{m}. \quad . \quad (53)$$

Before proceeding with the interpretation of this equation, let us consider the case $\mu=5$, and for this purpose let equation (52) be expressed in the form

$$R_{5\,;\,\nu}^{\nu}+R_5^{\nu}\Delta_\nu-\Delta_\nu R_5^{\nu}=0. \qquad (54)$$

Comparison with the case considered in Part II. p. 423, suggests that by analogy with the electromagnetic equations (R_5^{ν}) may be described as a current vector.

Reference to equation (53) supports this because the first term on the right-hand side can be interpreted as representing the Lorentz force, and R^{mn} is then proportional to the energy-momentum tensor of the field.

Following the notation adopted in the earlier paper, this tensor is written

$$E^{mn}=(h^2c^2/4\pi^2)R^{mn} \qquad (55)$$

and the current vector

$$J^n=(h^2c^2/4\pi^2)\alpha R_5^{\,n}. \qquad (56)$$

The interest in equation (53) lies chiefly in the occurrence of the term $(R_5^{\,m}\Gamma_0-\Gamma_0 R_5^{\,m})$. Associated with the term representing the Lorentz force on the current it must be interpreted as an additional action of the field on the current. This action is determined by the operator Γ_0, which is itself determined by the electromagnetic field in accordance with one of the relations (20). If equation (53) represents correctly the conditions prevailing in this field, any attempt at interpretation without the additional terms would appear to violate the principle of conservation of energy. A similar situation arises in connexion with equation (54) which, by the same method, may be shown to be expressible in the four-dimensional form

$$\left(\frac{\partial R_5^{\,n}}{\partial x^n}+R_5^{\,n}\Gamma_n-\Gamma_n R_5^{\,n}\right)=\Gamma_0 R_{55}-R_{55}\Gamma_0. \qquad (57)$$

The left-hand side of this equation corresponds to the expression for the divergence of a vector, and the vanishing of this expression is the equation of continuity. But it now appears that, although it is possible to express the equation of continuity in the usual form in a five-dimensional system, this is not the case in one of a four-dimensions, and again, if (57) represents the state of affairs in the field, the additional terms on the right must not be neglected. The suggestion is that we must look for some new source or sink of electric charge if $R_5^{\,n}$ is to be placed proportional to the electric current. The vanishing of the right-hand side of this equation would restore the position, and the present view of the position of the principle of conservation of electric charge in Physics might demand the adoption of some rule causing Γ_0 and R_{55} to commute.

But we are at present only at the threshold of theories of the nucleus and of its interaction with charged particles, and some modification of the principles underlying present theory is to be anticipated. In this connexion the results expressed in equations (53) and (57) may turn out to be suggestive.

The Expression for the Current Vector.

The value of the current vector R_5^n according to expressions (50) and (51) is

$$R_5^n = P_5^n = T^*_r T^{nr} + T_{5r} T^{*nr}$$

since $T_{55} = T^*_{55} = 0$.

After developing this expression in terms of the four-dimensional counterparts (A_{mn}) of (T_{mn}), it is found that

$$R_5^n = \sigma(\psi_r A^*_{nr} - \psi_r^* A_{nr}), \quad \ldots \ldots \quad (58)$$

where again the gravitational field is neglected.

The expression for A_{nr} is found in the usual way by means of equations (1)–(10) (Part II.), from which and from equations (30) and (33) it follows that

$$\left.\begin{aligned} A_{nr} &= \delta_n \psi_r - \delta_r \psi_n \\ A^*_{nr} &= \delta_n^* \psi_r^* - \delta_r^* \psi_n^* \end{aligned}\right\}. \quad \ldots \ldots \quad (59)$$

From these expressions and from equations (35) and (39) it follows that

$$R_5^n = \sigma \left\{ \psi_r \delta_n^* \psi_r^* - \psi_r^* \delta_n \psi_r + \frac{\alpha}{2\sigma} H_{pm}(A^*_{pm}\psi_n + A_{pm}\psi_n^*) + D_r(\psi_r^*\psi_n - \psi_r\psi_n^*) \right. . \quad \ldots \quad (60)$$

If there is no electromagnetic field this expression reduces to

$$\sigma \left\{ \psi_r \frac{\partial \psi_r^*}{\partial x^n} - \psi_r^* \frac{\partial \psi_r}{\partial x^n} + \frac{\partial}{\partial x^r}(\psi_r^* \psi_n - \psi_r \psi_n^*) \right\}, \quad \ldots \quad (61)$$

which is the same as the expression obtained in the treatment by means of unquantized equations except that in the present case the ψ's are operators.

Comparison with the equations of the electromagnetic theory helps in the interpretation of one of the components of this expression.

In applying Maxwell's equations to the theory of matter the polarization, P, and the intensity of magnetisation, M, are introduced by means of the relations

$$D = E + 4\pi P, \quad B = H + 4\pi M.$$

If D and H are eliminated from the field equations the following results are obtained :

$$\text{div } E = \rho - 4\pi \text{ div } P,$$

$$\text{curl } B = 4\pi \left(\frac{E}{c} + J + \frac{4\pi}{c} \dot{P} + \text{curl } M \right).$$

Thus a contribution to the current is obtained from a moment, M, and the moment, M, applicable to equation (61) has a component

$$\mathrm{M}_{nr}=\frac{h^2c^2}{4\pi^2}\alpha\sigma(\psi_r{}^*\psi_n-\psi_r\psi_n{}^*) \quad . \; . \; . \; . \; . \; . \; . \quad (62)$$

by equation (56).

If the field equations be expressed in Hamiltonian form and the concept of particles be introduced, this moment will be regarded as an intrinsic property of the particles.

From the point of view of the field the moment is expressed in terms of the quantities $(\psi_n, \psi_n{}^*)$ which by the manner of their introduction are field quantities corresponding in this theory to the components of electromagnetic potential (ϕ_n) in the electromagnetic theory. This form for the intrinsic moment was suggested by Proca$^{(1)}$ in its unquantized form and was adopted by Yukawa$^{(3)}$ in the operator form in his theory of the meson. It is now shown to be derivable in the operator form as a consequence of the principles adopted for the development of an operator field theory.

Note added on reading the proof, April 23rd, 1942.—The constant multiplier $h^2c^2/4\pi^2$ was introduced in equation (55) in order to make the results of this presentation of the field theory comparable with some other contributions to the subject. In order to compare it with a more recent paper by Kemmer$^{(4)}$ it is to be noted that his functions (ϕ_n) are equal to $(\kappa\psi_n)$, where $\kappa=2\pi\mu_0c/h$. In this paper the energy density is given by

$$\mathrm{E}_{44}=-2c\kappa\mathrm{R}_{44}$$

and the current density by

$$\mathrm{J}^n=-2c\kappa\alpha\mathrm{R}_5{}^n.$$

From this it follows, as in ii. p. 425, that a moment (M_{nr}) may be considered as contributing to the current where

$$\mathrm{M}_{nr}=-2\iota c\kappa^2\alpha(\psi_r{}^*\psi_n-\psi_r\psi_n{}^*).$$

If $\theta_n=\kappa\psi_n$ be introduced to avoid the use of ϕ_n, which may be confused with the electromagnetic potential, it follows that

$$\mathrm{M}_{nr}=2ie/m_0c(\theta_r\theta_n{}^*-\theta_r{}^*\theta_n).$$

From the relation between magnetic moment and spin it is suggested that the expression $2i(\theta_r\theta_n{}^*-\theta_r{}^*\theta_n)$ is a measure of spin for the meson field.

A point about the notation should be mentioned to avoid confusion. The quantity $(\mathrm{T}^{*\mu\nu})$ has been described as the complex conjugate of $(\mathrm{T}^{\mu\nu})$ This is not the case when one of the suffixes has the value 4. It is convenient to introduce the co-ordinate x^0 where $x^4=\iota x^0$, and in conformity with this $\psi^4=\iota\psi^0$, $T^{m4}=\iota T^{m0}$. In the case of the covariant relations $\psi_4=-\iota\psi_0$, $T_{m4}=-\iota T_{m0}$. The starred quantities are similarly related, *e. g.*, $T^{*m4}=\iota T^{*m0}$, so that although T^{m0} and T^{*m0} are conjugate complex quantities this is not the case for ψ^4, ψ^{*4} nor for T^{m4}, T^{*m4}.

References.

(1) Fisher and Flint, Proc. Roy. Soc. A, cxxvi. p. 645 (1929) ; Proca, J. Phys. Radium, vii. p. 347 (1936) ; and viii. p. 23 (1937).
(2) E. Schroedinger, *Sitz. Preuss. Akad. Wissenschaften*, xi. p. 105 (1932).
(3) Yukawa Sakata and Taketani, Proc. Phys. Math. Soc. Japan, xx. p. 319 (1940).
(4) N. Kemmer, Proc. Roy. Soc. A, clxvi. p. 124 (1938).

XXXIX. *Structure and Capacity of Electrical Double Layer.*

By J. J. Bikerman †.

[Received March 9, 1942.]

1. *The Fundamental Equation.*

Recent measurements of the capacity of the double layer at mercury-solution interfaces are so extensive and consistent that the time seems to be ripe for a new theoretical treatment of the subject.

The fundamental point of view remains that of Gouy (1) : within the diffuse double layer an equilibrium exists between external electric forces, which tend to separate positive from negative ions, and osmotic forces, which tend to equalize the distribution of every ionic species everywhere. In Gouy's theory ions are treated as electric charges having no volume and no other individual property. One example may suffice to show the difficulties of this approximation. Gouy's equations give for the concentration of monovalent cations near the mercury surface in a N solution at a potential of -1 volt the value 2.10^{17} N, and for the space available for one cation 8.10^{-39} c.c. These values are impossible, and clearly show the necessity of a stricter treatment.

The present paper takes into account the volume of ions and some of their properties. If the work done by electrostatic forces when moving a cation (or anion) from infinity to a distance x from the interface is ω_1 (or ω_2), the equation of Maxwell and Boltzmann gives for the concentration C_1' and C_2' of cations and anions at the distance x

$$C_1'=C_0e^{\omega_1/kT}\,; \quad C_2'=C_0e^{\omega_2/kT}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Here k is Boltzmann's constant and T the absolute temperature. For simplicity's sake the electrolyte is assumed to be a binary one, and C_0 is the common concentration (in gram-ions per c.c.) of cations or anions far away from the interface. In the Maxwell-Boltzmann equation the concentrations C_0, C_1', and C_2' are numbers of gram-ions per unit of volume *available to the ions.* If C, C_1, and C_2 are the respective concentrations per c.c. of the total volume, v_1 and v_2 the volumes occupied by a gram-ion of cation and anion respectively, and $2v=v_1+v_2$, equations (1) are transformed into

---

† Communicated by the Author.

$$C_1=Ce^{\omega_1/kT}\frac{1-C_1v_1-C_2v_2}{1-2Cv}, \quad \ldots \ldots \quad (2)$$

If C_2 is the analogous magnitude for anions, ν is the common valency of cations and anions, and F Faraday's constant, the free charge per c.c. of the solution in x cm. from the interface is

$$\rho=\nu FC_1-\nu FC_2. \quad \ldots \ldots \quad (3)$$

The second correction, briefly introduced two years ago (2), concerns the values of the work ω_1 and ω_2. An ion moving in an aqueous solution permanently exchanges places with water dipoles. If this movement occurs in an inhomogeneous electric field a displacement of dipoles generally requires a (positive or negative) work. Hence, the expression for the work done when transferring an ion from infinity to a short distance x from the mercury surface contains two terms, one for the change of position of the ionic charge, and another for that of the dipoles displaced by the ion. If ψ is the potential and $\frac{d\psi}{dx}$ the field intensity at a distance x from the mercury, the work $\omega_1=-\epsilon\psi\pm(\mu_1-\mu_0)\frac{d\psi}{dx}$; far from the interface both ψ and $\frac{d\psi}{dx}$ are equal to zero. Here ϵ is the charge on the cation, μ_1 is the dipole moment of the ion including its "hydration shell," *i. e.* taking into account the distortion of water molecules round the ion, and μ_0 is the sum of the dipole moments of the water molecules displaced by the ion and its water satellites. In the following computation the assumption is made that the difference $\mu_1-\mu_0$ is independent of ψ and concentration; more exact assumptions must be left for a future investigation. The sign before $\epsilon\psi$ is negative, as the electrostatic forces do negative work when a positive ion is lifted from $\psi=0$ to $\psi>0$. The sign before the term $(\mu_1-\mu_0)\frac{d\psi}{dx}$ is positive when $\frac{d\psi}{dx}$ is positive, *i. e.* when $\psi<0$. If $\left|\frac{d\psi}{dx}\right|$ is the absolute value of $\frac{d\psi}{dx}$,

$$\omega_1=-\epsilon\psi+(\mu_1-\mu_0)\left|\frac{d\psi}{dx}\right|. \quad \ldots \ldots \quad (4)$$

For anions in the analogous manner the equation

$$\omega_2=\epsilon\psi+(\mu_2-\mu_0')\left|\frac{d\psi}{dx}\right| \quad \ldots \ldots \quad (5)$$

is obtained; μ_2 and μ_0' are analogous to μ_1 and μ_0 respectively. If $\mu_1>\mu_0$, that is, if the "hydrated" cation has a larger dipole moment than the water it displaces, the accumulation of the cations in the diffuse double layer is denser than it would have been in the absence of the dipole effect. If $\mu_1 < \mu_0$, water dipoles are more strongly attracted to the region of the most intensive field than the cations, and the spreading of the diffuse double layer is increased.

This second correction can also be put in another form. The energy of a volume δv, in which the field intensity is $\frac{d\psi}{dx}$ and the dielectric constant D, is $\frac{D}{8\pi}\left(\frac{d\psi}{dx}\right)^2\delta v$. If the dielectric constant changes without a rearrangement of the electric charge, Maxwell's displacement $D\frac{d\psi}{dx}$ remains constant, and the energy of the volume can be written $\frac{\delta v}{8\pi D}\left(\frac{D d\psi}{dx}\right)^2$. If, therefore, a volume δv_1 containing a cation and its "hydration sphere" has the dielectric constant D_1 and an equal volume of undisturbed water the dielectric constant D_0, the energy change associated with the displacement of water by a cation is

$$\frac{\delta v_1}{8\pi}\left(\frac{1}{D_1}-\frac{1}{D_0}\right)\left(\frac{D\,d\psi}{dx}\right)^2=\vartheta_1\left(\frac{D\,d\psi}{dx}\right)^2 . \qquad (6)$$

For an anion it is

$$\frac{\delta v_2}{8\pi}\left(\frac{1}{D_2}-\frac{1}{D_0}\right)\left(\frac{D\,d\psi}{dx}\right)^2=\vartheta_2\left(\frac{D\,d\psi}{dx}\right)^2 . \qquad (7)$$

Hence

$$\omega_1=-\epsilon\psi-\vartheta_1\left(\frac{D\,d\psi}{dx}\right)^2, \qquad (8)$$

$$\omega_2=\epsilon\psi-\vartheta_2\left(\frac{D\,d\psi}{dx}\right)^2 . \qquad (9)$$

The polarization of an agglomeration of dipoles ceases to be proportional to the field intensity when the product $\mu\frac{d\psi}{dx}$ becomes larger than kT. For $kT=4.10^{-14}$ and $\mu=2.10^{-18}$ this condition is reached when $\frac{d\psi}{dx}>2.10^4$ electrostatic units. At potentials above, say, 0·5 volt from the electrocapillary maximum the field intensity at the interface appears to have the magnitude of 10^5 electrostatic units, therefore equations (4) and (5) represent a better approximation at high and equations (8) and (9) at low potentials.

Neither of the corrections suggested here is quite new. Stern (3) took into account the volume of ions by introducing into Gouy's theory the condition that only a limited number of ions could be accommodated in the monomolecular layer next to the mercury surface. The volume correction elsewhere was not considered, and the idea of special properties of monomolecular layers has lost much of its appeal since 1924.

A disadvantage of Stern's theory is that it contains too many unknown magnitudes which can be adjusted to suit the experimental data (see a similar remark by Audubert (4)). Two achievements of the theory were claimed by its supporters, namely, the explanation of the maximum of the electrokinetic potential ζ and of the difference between this potential and the total (transverse) potential difference. But the maximum of ζ disappears if ζ is calculated more correctly (5), and the difference between "transverse" and "tangential" potentials, in so far as it depends on the double layer, is probably due to roughness or swelling of solid surfaces (6).

The relative polarizability of solute and solvent was introduced by Gouy (7) and, in a quantitative form, by Frumkin (8) to interpret the effect of non-electrolytes on the electrocapillary curve. Butler (9) has published another version of this theory. The effect of ions on the polarization has not yet been considered.

No attempt is made here to calculate properly the effect of inter-ionic forces. They will be taken into account in so far only as activities will be substituted for concentrations when calculating the capacities at low potentials. The difference between activity and concentration appears, at high potentials, to be less important than the corrections put forward here.

Combining (3) with Poisson's equation $\frac{d^2\psi}{dx^2}=-\frac{4\pi\rho}{\mathrm{D}}$, the fundamental relation of the theory is obtained ;

$$\frac{d^2\psi}{dx^2}=-\frac{4\pi}{\mathrm{D}}\nu\mathrm{FC}(e^{\omega_1/k\mathrm{T}}-e^{\omega_2/k\mathrm{T}}).\frac{1-\mathrm{C}_1v_1-\mathrm{C}_2v_2}{1-2\mathrm{C}v}\quad . \quad . \quad . \quad . \quad (10)$$

ω_1 and ω_2 may be given either by (4) and (5) or by (8) and (9). If $\mathrm{C}v$, C_1v_1, C_2v_2 and either $(\mu_1-\mu_0)\frac{d\psi}{dx}$ and $(\mu_2-\mu_0')\frac{d\psi}{dx}$ or $\vartheta_1\left(\frac{\mathrm{D}\,d\psi}{dx}\right)^2$ and $\vartheta_2\left(\frac{\mathrm{D}\,d\psi}{dx}\right)^2$ are very small, equation (10) reduces itself to Gouy's equation :

$$\frac{d^2\psi}{dx^2}=-\frac{4\pi}{\mathrm{D}}\nu\mathrm{FC}(e^{-\varepsilon\psi/k\mathrm{T}}-e^{\varepsilon\psi/k\mathrm{T}}).$$

II. *High Potentials (Dipoles).*

When the potential difference ζ between $x=0$ and $x=\infty$ is high, the double layer contains so much more ions of one than of the other sign that, except very far from the boundary, equation (10) can be simplified to

$$\frac{d^2\psi}{dx^2}=-\frac{4\pi}{\mathrm{D}}\nu\mathrm{FC}e^{\omega_1/k\mathrm{T}}.\frac{1-\mathrm{C}_1v_1}{1-2\mathrm{C}v},\quad . \quad . \quad . \quad . \quad . \quad (11)$$

if the mercury is negatively, and to

$$\frac{d^2\psi}{dx^2}=+\frac{4\pi}{\mathrm{D}}\nu\mathrm{FC}e^{\omega_2/k\mathrm{T}}\cdot\frac{1-\mathrm{C}_2v_2}{1-2\mathrm{C}v}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

if it is positively charged.

(a) *Negative Branch.*

Along the negative branch $\psi<0$, $\frac{d\psi}{dx}>0$. Since

$$\mathrm{C}_1=\frac{\mathrm{C}e^{\omega_1/k\mathrm{T}}}{\mathrm{C}v_1e^{\omega_1/k\mathrm{T}}+1-2\mathrm{C}v},$$

equation (11) becomes

$$\frac{d^2\psi}{dx^2}=-\frac{4\pi\nu\mathrm{F}}{\mathrm{D}}\cdot\frac{\mathrm{C}e^{\omega_1/k\mathrm{T}}}{\mathrm{C}v_1e^{\omega_1/k\mathrm{T}}+1-2\mathrm{C}v}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

It is seen that the magnitude of $\frac{d^2\psi}{dx^2}$ (and of the charge density which, in virtue of the Poisson equation, is proportional to $\frac{d^2\psi}{dx^2}$) cannot rise indefinitely when ω_1 rises; it tends to a limit

$$\left(\frac{d^2\psi}{dx^2}\right)_{\max}=-\frac{4\pi\nu\mathrm{F}}{\mathrm{D}v_1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

The limiting state is approximately reached when $\mathrm{C}v_1e^{\omega_1/k\mathrm{T}}\gg1-2\mathrm{C}v$, that is,

$$e^{\omega_1/k\mathrm{T}}\gg\frac{1-2\mathrm{C}v}{\mathrm{C}v_1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

The value of $(1-2\mathrm{C}v)/\mathrm{C}v_1$ cannot be ascertained, but, fortunately, only its order of magnitude is required when estimating ω_1. In a N solution one cation is present for 55 molecules of water. The volume of a cation plus its " solvent envelope " cannot be less than that of a molecule of water. The highest value of $1/\mathrm{C}v_1$ is, therefore, for a N solution, 55; if the cation is strongly " hydrated " $1/\mathrm{C}v_1$ is much smaller. $\mathrm{C}v$ may be assumed equal to $\mathrm{C}v_1$. If $\frac{\omega_1}{k\mathrm{T}}>\log\mathrm{nat}\,150\approx5$, the inequality (15) is almost certainly satisfied. Equation (4) gives then

$$-\epsilon\psi+(\mu_1-\mu_0)\frac{d\psi}{dx}>5k\mathrm{T}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

If $\mu_1=u_0$ the saturation of the diffuse double layer occurs near $\psi=-0{\cdot}12$ volt, if the cations are monovalent. If $\mu_1>\mu_0$, both terms $-\epsilon\psi$ and $(\mu_1-\mu_0)\frac{d\psi}{dx}$ are positive and the condition (16) is reached at potentials below 0·12 volt. Both these cases are apparently unknown.

In all systems so far investigated μ_1 appears to be smaller than μ_0; the cations are less "polarizable" than the water they displace. Then $-\epsilon\psi$ is positive and $(\mu_1-\mu_0)\frac{d\psi}{dx}$ negative. If both are considerably larger than $5k\mathrm{T}$, the saturation of the double layer becomes noticeable when

$$e\psi=(\mu_1-\mu_0)\frac{d\psi}{dx}. \quad \ldots\ldots \quad (17)$$

When ζ increases the saturation takes place first in the immediate vicinity of the mercury-solution interface. The potential at the boundary is ζ, and the field intensity there may be denoted by $\left(\frac{d\psi}{dx}\right)_0$. Equation (17) becomes, then,

$$\epsilon\zeta=(\mu_1-\mu_0)\left(\frac{d\psi}{dx}\right)_0. \quad \ldots\ldots \quad (18)$$

From this equation the capacity of the double layer is easily calculated. Let the capacity per sq. cm. be K. It is σ/ζ, σ being the surface-charge density on mercury. Since

$$\sigma=-\frac{\mathrm{D}'}{4\pi}\left(\frac{d\psi}{dx}\right)_0, \quad \ldots\ldots \quad (19)$$

D′ being an average dielectric constant of the double layer,

$$\mathrm{K}=-\frac{\mathrm{D}'}{4\pi\zeta}\left(\frac{d\psi}{dx}\right)_0, \quad \ldots\ldots \quad (20)$$

and, introducing (18) †,

$$\mathrm{K}=-\frac{\mathrm{D}'\epsilon}{4\pi(\mu_1-\mu_0)}. \quad \ldots\ldots \quad (21)$$

The capacity per sq. cm. of a flat condenser is $\mathrm{D}'/4\pi\delta$, if δ is the thickness of the condenser. It is seen that for high potentials the "equivalent thickness of the double layer" is $l=\frac{\mu_0-\mu_1}{\epsilon}$; for monovalent ions it is the distance between the charges in the "resultant dipole."

When ζ increases further, the saturated part of the double layer becomes thicker. Let the thickness of the region of saturation be x_* cm. If ψ_* and $\left(\frac{d\psi}{dx}\right)_*$ are the potential and the field strength at the distance x_* from the interface, instead of equation (18) equation

$$\epsilon\psi_*=(\mu_1-\mu_0)\left(\frac{d\psi}{dx}\right)_*. \quad \ldots\ldots \quad (22)$$

---

† If $5k\mathrm{T}$ cannot be neglected compared with $\epsilon\zeta$, instead of (21) equation

$$\mathrm{K}=-\frac{\mathrm{D}'(\epsilon+5k\mathrm{T}/\zeta)}{4\pi(\mu_1-\mu_0)}. \quad \ldots\ldots \quad (21\,a)$$

results.

is obtained. The field intensity at $x=0$ is

$$\left(\frac{d\psi}{dx}\right)_0=\left(\frac{d\psi}{dx}\right)_*-\left(\frac{d^2\psi}{dx^2}\right)_{\max}x_*, \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

and the potential

$$\zeta=\psi_*-\left(\frac{d\psi}{dx}\right)_0 x_*-\frac{1}{2}\left(\frac{d^2\psi}{dx^2}\right)_{\max}x_*^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

From (22), (23) and (24) the relation follows

$$\frac{1}{\zeta}\left(\frac{d\psi}{dx}\right)_0=-\frac{1+\frac{1}{\zeta}\left(\frac{d^2\psi}{dx^2}\right)_{\max}x_*\left(\frac{x_*}{2}+l\right)}{x_*+l}. \quad . \quad . \quad . \quad . \quad . \quad (25)$$

When probable values for $\left(\frac{d^2\psi}{dx^2}\right)_{\max}$, see equation (14), and for l (see below) are introduced, it can be seen that the positive magnitude $-\frac{1}{\zeta}\left(\frac{d\psi}{dx}\right)_0$ increases very slightly when x_* increases, as long as the latter remains smaller than l.

Thus the theory requires the capacity to be nearly constant in a range of potentials above that given by the approximation (18). The approximate constancy of the capacity at high potentials has been observed many times †: see Gouy (10), Philpot (11), Craxford (11a), Vorsina and Frumkin (12), Frumkin (13) and Grahame (14), as well as references in these papers. To facilitate the comparison of theoretical and experimental results a figure from Frumkin's paper ‡ is reproduced here. It is seen that for N KCl the nearly constant value of the capacity is reached at about $-0{\cdot}4$ volts, reckoned from the maximum of the electrocapillary curve. The energy $-\epsilon\zeta$ corresponding to $\zeta=-0{\cdot}4$ volt is $16k\mathrm{T}$; formula (16) gives then $(\mu_1-\mu_0)\left(\frac{d\psi}{dx}\right)_0<11k\mathrm{T}$.

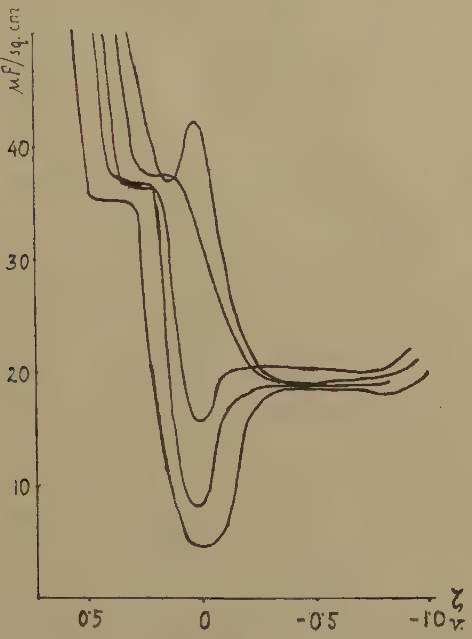
The ordinate of the level part of the K-ζ curve is given by equation (21). Since the expression for K contains two unknown magnitudes, namely D′ and $\mu_1-\mu_0$, it cannot be used for computing K, but from the experimental value of K the ratios $(\mu_1-\mu_0)/\mathrm{D}'$ and l/D' can be found. Since K for various KCl solutions is about 18 μF./sq. cm., *i. e.* about $1{\cdot}6.10^7$ cm.$^{-1}$, $(\mu_1-\mu_0)/\mathrm{D}'$ is $-2{\cdot}4.10^{-18}$ cm.$^{5/2}$ g.$^{1/2}$ sec.$^{-1}$ and l/D' is 5.10^{-9} cm. The dielectric constant cannot be constant across the whole thickness of the double layer, and its value to be inserted in equation (21) is a complicated function of all the D values represented in the layer. It is certain, however, that it is above 2 and much less than 80. For $D'=2$, 15, and 80 the value of μ_0-u_1 would be 5.10^{-18}, 36.10^{-18}, and 192.10^{-18}, and the value of l 10^{-8}, $7{\cdot}5.10^{-8}$, and 40.10^{-8} cm. The lower figures for $\mu_0-\mu_1$ and l are quite of the expected order †.

---

† The capacities recorded in the literature are mostly "differential capacities," whereas K calculated here is the "static capacity." Since $\sigma=\mathrm{K}\zeta$, $d\sigma/d\zeta=\mathrm{K}+\zeta.d\mathrm{K}/d\zeta$; this equation gives the differential capacity $d\sigma/d\zeta$ as a function of K and ζ. The difference between K and $d\sigma/d\zeta$ is here not large enough to affect the comparison made in the text.

‡ Trans. Farad. Soc. xxxvi. p. 124.

The capacity as expressed by equation (21) is independent of the concentration of the electrolyte‡. Experimental results are in agreement with this deduction from the theory. The potential ζ at which the level part of the K-ζ curve commences should, on the contrary, depend on concentration. The value of $(1-2Cv)/Cv_1$ was approximately estimated for a N solution. If the solution is weaker, say 10^{-4} N, the values of Cv and Cv_1 are 10^4 times as low (v and v_1 may, of course, depend on C in concentrated solutions, but this effect is neglected for the time



Capacity of a mercury electrode (after **Frumkin**). The curves refer to N KCl (top), 0·1 N KCl, 0·01 N KCl, 0·001 N KCl, and 0·0001 N HCl (bottom).

---

† If equation (21*a*) instead of (21) is used, all the values of $\mu_0-\mu_1$, and of l, would be reduced by about 25 per cent.

‡ A more exact relation is $K=\frac{D'}{4\pi}.\frac{\epsilon'}{\mu_0-\mu_1}\left(1-\frac{nkT}{-\epsilon\zeta}\right)$, n being a function of Cv and Cv_1.

being), but the log nat of $(1-2Cv)/Cv_1$ is only increased by about 9, so that $\frac{\omega_1}{kT}$ must be over, say, 14 in order to cause a nearly constant value of K. In very dilute solutions the term $(\mu_1-\mu_0)\left(\frac{d\psi}{dx}\right)_0$ is less important than in concentrated ones (since the capacity in dilute solutions is less) and the condition (15) is almost equivalent to $-\epsilon\zeta>14kT$, that is, $-\zeta>0{\cdot}35$ volt. This agrees with Frumkin's curve for 10^{-4} N HCl, which shows a bend at about $-0{\cdot}35$ volt.

(b) *Positive Branch.*

The treatment of the anionic layer, *i. e.* of the positive branch of the electrocapillary curve $\left(\psi>0, \frac{d\psi}{dx}<0\right)$ is perfectly analogous to that of the cationic layer. The nearly constant capacity K of the layer consisting of Cl ions is about 37 μF./sq. cm., *i. e.* $3{\cdot}3 \,.\, 10^7$ cm.$^{-1}$. That gives for D′ 2, 15, and 80 values of $\mu_0'-\mu_2$, equal to $2{\cdot}5 \,.\, 10^{-18}$, $19 \,.\, 10^{-18}$, and $101 \,.\, 10^{-18}$ The shift of the level part of the K-ζ curve to higher potentials when concentration decreases is clearly seen in Frumkin's curves for 0·1 to 0·001 N KCl The value of the capacity is again independent of the concentration, in accord with equation (21).

From expression (16) it follows that a saturation of the diffuse double layer takes place at lower potentials if the resultant dipole moment is smaller. Since the values of $\mu_0'-\mu_2$ calculated above are smaller than those of $\mu_0-\mu_1$, the nearly constant value of K should be reached for anionic layer at lower potentials than for cationic layer. This effect is also visible in the figure.

III. *High Potentials.* (*Dielectric Constant.*)

When into expression (15) ω_1 from (8) is introduced, instead of (21) the relation

$$K=\frac{D'}{4\pi D}\left(-\frac{\epsilon}{\vartheta_1\zeta}\right)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

results. Repetition of the arguments leading to equation (25) shows that this capacity as well varies but little when ζ rises above the value determined by (15). From (26) the order of magnitude of ϑ_1 can be calculated for probable values of the ratio D′/D, but no data are given here since the "dipole approximation" is, at high potentials, more accurate than the "dielectric constant approximation."

IV. *Low Potentials.* (*Dipoles.*)

When ζ is low, equation (10) can be simplified to

$$\frac{d^2\psi}{dx^2}=-\frac{4\pi\nu FC}{D}\cdot\frac{\omega_1-\omega_2}{kT}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

Equations (4) and (5) give then

$$\frac{d^2\psi}{dx^2}=-\frac{4\pi\nu\mathrm{FC}}{\mathrm{D}k\mathrm{T}}\left[(-\epsilon-\epsilon)\psi+(\mu_1-\mu_0-\mu_2+\mu_0{}')\left|\frac{d\psi}{dx}\right|\right]. \quad . \quad (28)$$

(a) *Negative Branch.*

Introducing

$$\frac{4\pi\nu\mathrm{FC}}{\mathrm{D}k\mathrm{T}}=\beta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

and $\mu_1-\mu_0-\mu_2+\mu_0{}'=-\mu$, we get

$$\frac{d^2\psi}{dx^2}=2\beta\epsilon\psi+\beta\mu\frac{d\psi}{dx}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

For N aqueous solutions at 20° β is about $1{\cdot}1 \,.\, 10^{24}$ cm.$^{-7/2}$ g.$^{-1/2}$ sec.; if activity is substituted for concentration β becomes (for N KCl) $7\,.\,10^{23}$.

If D is assumed to be practically independent of x, equation (30) is satisfied by setting

$$\psi=\zeta e^{-kx}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

ζ is negative. k is positive and given by the relation

$$k=-\frac{\beta\mu}{2}+\left(\frac{\beta^2\mu^2}{4}+2\beta\epsilon\right)^{1/2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

From (31)

$$\left(\frac{d\psi}{dx}\right)_0=-\zeta k \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

is obtained; hence the capacity per sq. cm. is

$$\mathrm{K}=-\frac{\mathrm{D}'}{4\pi\zeta}\left(\frac{d\psi}{dx}\right)_0=\frac{\mathrm{D}'k}{4\pi}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

Evidently k^{-1} is the "thickness of the equivalent double layer."

When $\mu=0$, or the concentration is very small, the product $\beta^2\mu^2$ is much smaller than $8\beta\epsilon$ and, approximately,

$$k=\sqrt{2\beta\epsilon}=\left(\frac{8\pi\nu\mathrm{F}_\varepsilon\mathrm{C}}{\mathrm{D}k\mathrm{T}}\right)^{1/2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

This expression is identical with that given by Gouy for the "inverse thickness of the equivalent double layer" or with κ of Debye's theory †. In agreement with this result Frumkin's measurements show that the capacity of dilute solutions (0·001 N KCl and 0·0001 N HCl) is approximately equal to that predicted by Gouy. Actually it is somewhat larger, *e. g.* for 0·0001 N HCl 3·9 instead of 2·4 μF./sq. cm. This discrepancy probably can be accounted for by the difference between "static capacity" K and "differential capacity" $d\sigma/d\zeta$ (see footnote on p. 390). The capacities in dilute solutions near $\zeta=0$ increase with increasing absolute values of ζ; the term $\zeta\frac{d\mathrm{K}}{d\zeta}$ is positive and the differential capacity $\mathrm{K}+\zeta d\mathrm{K}/d\zeta$ measured by Frumkin is larger than the static capacity K calculated by Gouy.

---

† The unusual form employed here is due to referring ϵ and kT to single ions and F and C to gram-ions.

When μ is small, the capacity in concentrated solutions should also be near that calculated by Gouy, *i. e.* very high. Activities instead of concentrations have to be employed here. Equations (34) and (35) give, *e. g.*, for N KCl, whose activity coefficient is 0·6, if D in (35) is equal to D′ in (34), for D=25 and 80 capacities of 100 and 180 μF./sq. cm. That is the reason for the maximum of capacity in concentrated solutions near the electrocapillary maximum. The "hump" on the K-ζ curve was first observed by Gouy (10) and found again by Frumkin (13) and by Grahame; no explanation of its existence has been advanced before.

If $\mu\neq 0$, it appears to be usually positive, *i. e.* the resultant moment of the anion is larger than that of the cation; that was the reason for substituting $-\mu$ (as distinct from μ) for the expression $\mu_1-\mu_0-\mu_2+\mu_0'$ in equation (28). If μ is very large, the product $\beta^2\mu^2$ is much larger than $8\beta\epsilon$, and instead of (32) $k=2\epsilon/\mu$ may be written; the capacity becomes

$$\mathrm{K}=\frac{\mathrm{D}'\epsilon}{2\pi\mu}. \qquad (36)$$

In the limiting case of high μ values the capacity, like that given by equation (21), is a function of the dielectric constant and a resultant dipole moment, but independent of concentration. This limit is only approached when (for N solutions) μ exceeds 10^{-16} (100 debyes), *i. e.* when an ion and its "hydration shell" displace some 50 water molecules. For $\mu=3.10^{-16}$ equation (36) gives $\mathrm{K}=22\mu$F./sq. cm., if $\mathrm{D}'=80$.

(b) *Positive Branch.*

If μ is positive, what appears to be the usual case, the dipole effect increases the capacity along the positive branch of the electrocapillary curve near its maximum. For k, instead of equation (32) the equation

$$k=\frac{\beta\mu}{2}+\left(\frac{\beta^2\mu^2}{4}+2\beta\epsilon_1\right)^{1/2} \qquad (37)$$

results. The cause of this increase of capacity is that, although the concentration of anions in the double layer is less than it would have been in the absence of the dipole effect, the difference between it and the concentration of cations is larger.

V. *Low Potentials.* (*Dielectric Constant.*)

Equation (27) gives with (8), (9), and (29)

$$\frac{d^2\psi}{dx^2}=2\beta\epsilon\psi-\beta(\vartheta_1-\vartheta_2)\mathrm{D}^2\left(\frac{d\psi}{dx}\right)^2. \qquad (38)$$

Hence the capacity per sq. cm.

$$K = -\frac{D'}{4\pi\zeta}\left(\frac{d\psi}{dx}\right)_0 = -\frac{D'\epsilon^{\frac{1}{2}}}{4\pi\zeta\beta^{\frac{1}{2}}D^2(\vartheta_1-\vartheta_2)}[e^{2\beta(\vartheta_1-\vartheta_2)D^2\zeta}-2\beta(\vartheta_1-\vartheta_2)D^2\zeta-1]^{\frac{1}{2}}, \quad . \; . \; . \quad (39)$$

the minus sign being valid for negative potentials. In integrating (38) the dielectric constant D was again treated as a constant. At low potentials the error thus introduced cannot be large, and it may also be assumed that the ratio D′/D is not far from unity.

When the product $\beta(\vartheta_1-\vartheta_2)D^2\zeta$ is very small, equation (35) is obtained again. Hence, at $\zeta=0$, the capacity is, except for the introduction of the activity coefficient, given by Gouy's equation and is very high at high concentrations. Thus the "hump" of the K-ζ curve is found also by the "dielectric constant treatment".

When the product $C(\vartheta_1-\vartheta_2)$ or $\beta(\vartheta_1-\vartheta_2)$ is very large, the capacity is nearly equal to

$$K = \frac{D'\epsilon^{1/2}}{2\sqrt{2\pi}|\zeta|^{1/2}(\vartheta_1-\vartheta_2)^{1/2}D} \quad . \; . \; . \; . \; . \; . \; . \quad (40)$$

already at relatively small ζ values. For instance, for $-\beta(\vartheta_1-\vartheta_2)D^2\zeta$ to be 5 at $\zeta=-0{\cdot}025$ volts in a N solution of a dielectric constant 80, $(\vartheta_1-\vartheta_2)$ has to be $1{\cdot}4 \,.\, 10^{-23}$ c.c., if the activity coefficient be 0·6.

If $\vartheta_1-\vartheta_2$ is $1{\cdot}4 \,.\, 10^{-23}$, equation (40) with $D'=D$ gives $K=80$ μF/sq.cm. The capacity of mercury surface in N KCl at $\zeta=-0{\cdot}025$ volt, according to Frumkin, is 40, but it is the differential capacity $d\sigma/d\zeta$. In the validity range of equation (40) the difference between K and $K+\zeta dK/d\zeta$ which could be neglected along the level part of the K-ζ curve is very noticeable. Indeed, a differentiation of (40) shows that $K+\zeta.dK/d\zeta=K/2$. As $d\sigma/d\zeta=40$ μF./sq.cm., $K=80$ μF./sq.cm. The experimental value of $d\sigma/d\zeta$ gives, therefore, $\vartheta_1-\vartheta_2=1{\cdot}4 \,.\, 10^{-23}$ c.c. Setting in equations (6) and (7), $D_1=2$ and $D_2=4$, $\vartheta_1-\vartheta_2$ becomes approximately $0{\cdot}01\,\delta v$, if $\delta v=\delta v_1=\delta v_2$. Hence, an average ion of KCl with its "hydration shell" appears to displace over 10^{-21} c.c. of water. This volume is too large, since a N solution contains $1{\cdot}2 \,.\, 10^{24}$ ions per litre, and the volume immobilized would be over 1·2 litre, but the uncertainty of the values of D, D′, D_1, and D_2 is so great that no attempt is made here to calculate δv with more precision. The volume of water immobilized by a monovalent ion according to the calculation by Sack (15) (see also Mason and Shutt (16)) is $7 \,.\, 10^{-21}$ c.c

Equation (39) requires K to drop from, say, 120–180 μF./sq. cm. at $\zeta=0$ to 80 at $-0{\cdot}025$ volt. This peak has not been detected so far.

Along the *positive branch* of the curve, when, as usual, ϑ_1–ϑ_2 is positive, equation (39) shows that the capacity in the positive neighbourhood of the electrocapillary maximum is higher than on the negative side. At $\zeta=0$ the K-ζ curve is inclined to the abscissa instead of being parallel to it, and the maximum of the K-ζ curve is observed at a lower applied voltage than that of the electrocapillary curve. This effect can be seen, for instance, in Gouy's data (10) for magnesium sulphate and on Frumkin's and Grahame's curves.

When $(\vartheta_1-\vartheta_2)$ or μ is large, the nearly constant value of K at high negative potentials is almost certainly smaller than the nearly constant capacity at high positive potentials, see (21); it may be expected, therefore, that the maximum of the K-ζ curve will be the higher and/or occur at the higher ζ values the greater is the difference between the capacities of the cationic and the anionic layer. An indication of this effect is apparent in Grahame's curves for N NaCl, N H_2SO_4, and N $NaNO_3$.

Since anions are less easily displaced by water from the boundary layer than cations are, a saturation of the double layer by anions takes place at a lower potential than along the negative branch. The transition from equation (27) to (15) is more rapid than in the cationic layer; this is best seen on Frumkin's curves for KCl.

VI. *Summary.*

A theory of the structure of the diffuse double layer is advanced. It is based on the consideration of the dipole properties or dielectric constants of ions and solvent. In an inhomogeneous field substances having a higher dipole moment per unit of volume (or a higher dielectric constant) tend to accumulate in places of the most intensive field at the expense of substances having lower moments. Therefore, in the competition between ions and solvent molecules for positions near the interface, their dipole moments (or dielectric constants) are decisive.

The theory is applied to the capacity of mercury-solution interfaces. For low potentials and low concentrations Gouy's values are confirmed, and for low potentials and high concentrations they are lowered by inter-ionic forces. At high potentials the thickness of the equivalent double layer is independent of the concentration and almost independent of the potential; it is equal to the distance between the charges in the resultant dipole. The theory accounts for the general shape of capacity-potential curves, including the maximum of capacity at low potentials in concentrated solutions, the position of this maximum, and the nearly constant capacity at high potentials.

References.

(1) G. Gouy, *J. de physique* (4), ix. p. 457 (1910).
(2) J. J. Bikerman, Trans. Farad. Soc. xxxvi. p. 718 (1940).

(3) O. Stern, *Z. f. Elektrochem.* xxx. p. 508 (1924).
(4) R. Audubert, Trans. Farad. Soc. xxxvi. p. 144 (1940).
(5) J. J. Bikerman, *Z. physik. Chem.* A, clxxi. p. 209 (1934) ; Trans. Farad. Soc. xxxvi. p. 154 (1940).
(6) J. J. Bikerman, J. Chem. Physics, ix. p. 880 (1941), and in the press.
(7) G. Gouy, *Ann. de physique* (9), vii. pp. 129, 175 (1917).
(8) A. Frumkin, *Z. f. Physik*, xxxv. p. 792 (1926).
(9) J. A. V. Butler, Proc. Roy. Soc. A, cxxii. p. 399 (1929) ; 'Electrocapillarity,' p. 82. (London, 1940.)
(10) G. Gouy, *Ann. de chim. et de phys.* (7) xxix. pp. 159, 230 (1903).
(11) L. St. J. Philpot, Phil. Mag. (7), xiii. p. 775 (1932).
(11*a*) S. R. Craxford, Thesis. (Oxford, 1936.)
(12) M. A. Vorsina and A. N. Frumkin, *C. R. Acad. Sci., U.S.S.R.*, xxiv. p. 918 (1939).
(13) A. N. Frumkin, Trans. Farad. Soc., xxxvi. p. 117 (1940).
(14) D. C. Grahame, J. Amer. Chem. Soc. lxiii. p. 1207 (1941).
(15) H. Sack, *Phys. Z.* xxvii. p. 206 (1926) ; xxviii. p. 199 (1927).
(16) W. A. Mason and W. J. Shutt, Proc. Roy. Soc. A, clxxv. p. 234 (1940).

XL. *Dimensions.*

By NORMAN CAMPBELL, Sc.D., F.Inst.P.*

[Received March 3, 1942.]

RECENT papers in this journal suggest the following remarks. Writers on "dimensions" usually ignore almost the whole of the vast literature of the subject. The reason cannot be that they are taking for granted established doctrines; for it is notorious that almost every proposition about dimensions, asserted by one writer, has been denied by a writer of equal competence. The reason is probably precisely the opposite: anyone who attempted to summarize the literature would become involved in so many controversies that he could get no further. But the absence of background makes it impossible to ascertain the significance of any particular contribution.

Now there is one doctrine on which all writers seem to be agreed. The assignment of dimensions is a means to an end, not an end in itself; the differences concern the end. If, then, all writers would state explicitly what end they have in view, and how that end is served by the considerations they present, their writings would become much more intelligible.

The following list of ends hitherto proposed is suggested; of course, anyone is entitled to add to it, but he should indicate briefly how the new end is related to the old.

(1) The avoidance of algebraic errors and of assigning undue importance to numerical coincidences. (Comparison of dimensions on either side of an equation.)

(2) The statement of laws in their most compact and general form. (Use of zero-dimensional magnitudes.)

(3) Elucidation of the best (perhaps the only adequate) form in which to express the results of measurement. (Choice of "units" and standards.)

(4) Deduction of numerical laws with the minimum of mathematical analysis. (Dimensional analysis.)

(5) Identification of ultimate theories. (*E.g.*, by the doctrine that all magnitudes have ultimate dimensions in mass, length and time.)

---

* Communicated by the Author.

The Editors do not hold themselves responsible for the views expressed by their correspondents.]

XLI. *Contact Potentials.*—Part I. *General Principles.*

By J. Alan Chalmers, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham *.

[Received February 4, 1942.]

1. *Introduction.*

At the end of the last century, there was a considerable controversy on the subject of "contact potentials," the main exponents of the two rival theories being Kelvin (1), with a "contact" or "physical" theory, and Lodge (2), with a "chemical" theory. The controversy died down without ever being finally settled and, as a result, the ideas of both theories still survive, sometimes in the same discussion, with consequent ambiguity as to the meaning of a "contact potential," and occasionally entirely false conclusions have been reached, *e. g.* when Hume-Rothery (3), discussing surface layers, states that "the contacts . . . are no longer of a simple physical nature."

It has been realized by many workers in electro-chemistry, *e. g.* Lange and Nagel (4), that the Kelvin theory, if applied consistently, can account for the phenomena satisfactorily, but it does not seem to have been realized that the Lodge theory, slightly modified, and with different definitions of "contact potentials," can equally well describe the facts, and that this point of view may be as good as, and in some respects better than, the account on the Kelvin theory.

It seems that no previous attempt has been made to give any complete account of the theory of "contact potentials," bringing both the older theories under consideration; and it seems that such an account should go far towards removing the ambiguities and apparent contradictions that occur in the subject.

It is necessary to devote some space to the statement of clear definitions and to the discussion of general principles before going on to the interpretation of experimental facts in terms of these principles. The general principles and general deductions from them will form the subject of Part I., while later parts will deal with experimental phenomena and their explanation.

---

* Communicated by the Author.

2. *Potential Differences.*

In the course of this work, we shall be continually referring to "potential differences" or, more shortly, "potentials," and it is necessary to have a clear idea of what we mean by a potential difference.

In electrostatic theory, a potential difference between two points is defined as the work done in transferring a unit charge from one point to the other. Such a definition is inadequate from the point of view of contact potentials without further consideration of the terms "work done" and "unit charge."

In purely electrostatic phenomena, there are no heat effects, and considerations of "isothermal" or "adiabatic," or of "constant pressure" or "constant volume," do not arise; but when, as in the phenomena of contact potentials, we discuss the transfer of particles from one medium to another, then these conditions do matter. Since it is possible for heat to enter or leave the system in any ordinary case of contact potentials that may be discussed, and since we shall not be directly concerned with changes of temperature, we can see that the conditions are isothermal. Since we shall only consider "virtual" transfers of charge, the charge densities remaining constant, the conditions are those of constant pressure rather than constant volume. The question of the conditions of constant temperature and pressure is exactly the same as in the case of the thermodynamic treatment of equilibrium between two phases and the change of state (*cf*. Guggenheim (5)).

If the forces opposing the motion are purely electrostatic, then there is no ambiguity in regard to the work done in moving a unit charge; but if, as occurs in the phenomena of contact potentials, there are forces which can broadly be described as "chemical," then the nature of the carrier of the charge does influence the work done in its motion.

Thermodynamics has provided functions which can be used in the discussion of contact potentials in the form of "thermodynamic potential" and "chemical potential" (*cf*. Guggenheim (5)). When there is equilibrium, under conditions of constant temperature and pressure, at a boundary, then the transfer of a particle across the boundary causes no change in the thermodynamic potential of the system, and so the chemical potential (or, more exactly, if the particle is charged, the "electrochemical potential") of the moving particle is the same on either side of the boundary. If one particle be removed from a phase, under conditions of constant temperature and constant pressure, then the change in the thermodynamic potential of the whole system is equal to the chemical potential of the particle within the phase $\left[\text{for } \mu = -\left(\frac{\partial G}{\partial n}\right)_{p,T}\right.$, using Guggenheim's notation$\left.\right]$.

We can define a potential difference between two points as the change in thermodynamic potential of the system due to the " virtual " transfer of a charge from one point to the other, bringing in, where necessary, the effect of changing from one carrier of charge to another. When there is no change of carrier, the potential difference is then the difference in the chemical potential of the carrier at the two points. It can easily be seen that such a definition of potential difference agrees with the electrostatic definition in the cases where the forces are purely electrostatic ; and it can be seen that the " contact potential difference between two substances " is ambiguous and requires further specification before its meaning is clear. so that there may be more than one " contact potential difference " at a boundary.

3. *Phases and Phase Boundaries.*

The phenomena of contact potentials are concerned with boundaries between different substances, and we can divide any system of substances into a number of *phases*, each one of which is homogeneous in itself. From the point of view of electrical phenomena, the important feature of a phase is the way in which it can carry an electric current, and we can divide phases into three main groups which can be broadly described as metals, electrolytes, and dielectrics.

Metals can be distinguished by the fact that the only way in which electricity can be carried is by electrons, the positively charged metal atoms remaining fixed in position. Liquid metals are practically the same, for the metal atoms only carry a very small proportion of the current. Other solid substances *e,. g.* the " semi-metallic " conductors which carry current by means of electrons, must also be included in the general category of " metals," and so must the " electron gas " in a vacuum that is considered in discussions on thermionic emission.

. In electrolytes the current is carried by ions of both signs, the ions being much heavier than electrons. In simple electrolytes there will only be one type of ion of each sign, but, if there are mixed electrolytes, we shall get different ions present. Included with electrolytes must be ionized gases and fused salts, for these conduct electricity by means of ions.

Dielectrics cannot conduct electricity, but there may be electric charges at their boundaries and electric forces across them. It is often necessary to discuss the effect of moving an imaginary charge across a dielectric, and in such cases it does not matter what type of carrier is imagined to carry the charge, the work done depending on the magnitude of the charge, not on the nature of the carrier.

At a boundary of two phases, it is necessary to discuss what carrier (ion or electron) takes the current across the boundary ; this can be inferred from the actual changes in the contents of the phases when a small current is imagined to pass. It is clear that different types of phase boundary will have different types of carrier across the boundary.

At a boundary of two metals, it is clear that any current must be carried by electrons, a simple case.

At a boundary of two electrolytes the matter is complicated, for there are at least four carriers of charge, and these may be all different, in which case all may carry a fraction of the current, depending on their relative numbers and velocities ; thus it is not possible to consider a single carrier, but, for some purposes, we can consider all the carriers grouped together, in proportion to the fraction of the current carried by each, to give a sort of " average " carrier ; at some electrolyte boundaries, *e. g.* when one of the " electrolytes " is an ionized gas, some of the carriers cannot cross the boundary, and the actual transfer of charge across may be due to electrons which discharge ions on both sides of the boundary.

Metal-electrolyte boundaries are of two kinds, which can be called " active " and " inert " according as to whether the metal ions can or can not go into solution. In the active case, it is the metal ion which carries the current across the boundary, but in the inert case it can only be an electron. In more complicated cases, the metal of an ion in the solution may differ from the metal electrode, and this metal may be deposited, so that the current may be carried by two metals or one metal and an electron.

4. *Surface Layers.*

In most, if not all, cases of boundaries between two phases, there is no clear-cut boundary between two pure and homogeneous phases ; but there is usually found a surface layer on one or both phases, differing in properties from the pure phase. It is probably never really correct to speak of the surface layer as a single separate phase, for the surface layer is not itself homogeneous, but varies in its properties through its thickness. But it is the properties of the surface layer at the interphase boundary (usually the surface of slip in relative motion) that matter in discussions of contact potentials, and it seems that there is no significant error in considering a surface layer to have the properties which it has at the boundary, and to consider a sharp boundary to exist between this surface layer and the underlying phase.

A surface layer is usually conveniently considered to have the same carrier or carriers of charge as exist in the underlying phase. The types of surface layer that occur can be divided into two classes according as to whether the layer is due to the phase on the other side of the boundary or not ; for example, mercury in a solution of potassium iodide carries a surface layer of iodine atoms or iodide ions ; while a metal in air or vacuum always has a surface layer unless specially cleaned.

5. *Equilibrium at Phase Boundaries.*

The conditions of equilibrium at a phase boundary are simplest if there is only one carrier of charge across the boundary, and we shall consider such a case first. When two phases are brought into contact, there may be a tendency for particles to move from one phase into the other, if the chemical potential of the carrier differs in the two phases. The passage of the carrier causes electric charges to appear on the two phases, and these electric charges alter the chemical potentials of the carrier in the two phases ; the transfer of charge continues until the chemical potentials are the same on both sides of the boundary, when there will be equilibrium. It is this equilibrium state which is discussed, and it is these electric charges which appear in the phenomena of contact potentials.

If charge can be carried across the boundary by more than one carrier, then it would only be by chance that the chemical potentials of each carrier could be the same on both sides of the boundary, so that, in general, there could not be true equilibrium, but a state could be reached in which the transfer of charge across the boundary is the same in either direction, giving what can be termed " merely electrical equilibrium " ; it is clear that thermodynamic reasoning cannot be applied in such a case in the same way as for true equilibrium.

6. *Measured Contact Potential Differences.*

A potential difference can be measured electrically either by the current which flows in a closed circuit or else by the work done in carrying a charge across a dielectric. It can be seen that in neither case is it possible to obtain a measurement of a contact potential difference at one boundary alone; a complete circuit containing one boundary of a pair of conductors must also contain either another boundary of the same conductors, which annuls the effect of the first (unless there are temperature differences, in which case neither conductor can be considered a homogeneous phase), or else some third substance, which gives at least two other boundaries in the circuit. When there is a dielectric present, there are at least the two dielectric-conductor boundaries as well as the boundary of the two conductors ; or, if both sides of the dielectric consist of the same conductor, these boundaries annul and the conditions are as in the complete circuit.

It is usually preferable to consider a system which is not giving a continuous current, although the conditions must be those of a " virtual " current ; in actual fact, most of the measurements of potential difference are carried out by methods which do not involve an actual flow of current, *e.g.* the use of a potentiometer for the measurement of the potential difference due to a chemical cell. So, when discussing a system which consists entirely of conductors, it is considered either as in a potentiometer circuit or to have a gap in one of the conductors, with a dielectric, so that the potential difference across this dielectric is measured, the two conductor-dielectric boundaries cancelling out in their effects.

It is only possible to obtain actual measurements of potential differences when there is a system containing at least three boundaries, and it is only possible to divide up this measured potential difference into parts corresponding to the separate boundaries if we make use of some theoretical argument as to how this is to be done.

For any two defined points P and Q in a system, it is possible to imagine, theoretically, the change in the thermodynamic potential of the system when a specified carrier of charge moves from P to Q ; it is not possible to measure this in practice because the carrier would have to get to P and from Q. We can define this potential difference as K_{PQ}, and it is possible to extend the idea of potential difference to include the cases where there are different carriers of charge at P and at Q by considering the difference of chemical potential of two carriers at the same point to be a part of the potential difference. We shall leave until later the discussion of units and sign conventions for potential differences.

If we take a charge right round a complete circuit, finishing at the same point from which we started, the total change in thermodynamic potential must be equal to the total measured potential difference, and so we get

$$L_{PQR\ldots OP}=K_{PQ}+K_{QR}+\ldots+K_{OP},$$

where $L_{PQR\ldots OP}$ is the measured potential difference and K_{PQ} etc. are the potential differences between points P and Q etc. as defined above.

If the phases are denoted by 1, 2, 3 etc., we can write the measured potential difference as $L_{123\ldots n1}$, and this is equal to $K_{12}+K_{23}+\ldots+K_{n1}$, where K_{12} represents the potential difference between some point in phase 1 and some point in phase 2, while K_{23} represents the potential difference between the same point in phase 2 and some point in phase 3, and so on, there being only one point in each phase to be considered. The actual values of K_{12} etc. depend on which points are used for the definitions of K_{12}, but the result that $L_{123\ldots n1}=K_{12}+K_{23}+\ldots K_{n1}$ holds so long as one point only is used in each phase.

7. *Systems of Potential Differences.*

As we have seen, we can only measure the potential difference of a system which contains at least three interphase boundaries. In this section we shall see that any more complex system can be considered as the sum of systems of three phases.

Since

$$L_{123\ldots n1}=K_{12}+K_{23}+\ldots K_{n1},$$

we can put

$$\begin{aligned}L_{123\ldots n1}&=K_{12}+K_{23}+K_{31}-K_{31}+\ldots K_{n1}\\&=L_{1231}-K_{31}+K_{34}+K_{41}-K_{44}+\ldots K_{n1}\\&=L_{1231}+L_{1341}-K_{41}+\ldots K_{n1}\\&=L_{1231}+L_{1341}+L_{1451}+\ldots L_{1(n-1)n1},\end{aligned}$$

so that the measured potential difference of the complete system has been divided into measured potential differences of systems of three phases.

The potential difference of a system can be divided up into a number of different ways for

$$L_{123\ldots n1}=L_{12k1}+L_{23k2}+\ldots L_{n1kn},$$

whatever phase k may be.

As an example of the way in which this result can be used, let us consider a system of three phases, 1, 2, and 3 with a surface layer 1′ between phases 1 and 2.

Then $L_{11'231}=L_{11'21}+L_{1231}$, of which the first represents the effect of the surface layer and the second that of the pure phase.

The systems of three phases that can be considered can be divided into ten groups by considering metals, electrolytes and dielectrics; of these the most important are:

I. Three metals.
II. Two metals and a dielectric (Volta effect).
III. Two metals and an electrolyte (Voltaic cell).
IV. One metal, one electrolyte, and a dielectric (Metal-electrolyte Volta effect).
V. One metal and two electrolytes (Concentration cell, etc.).

8. *Systems of Zero Potential Difference.*

There are certain systems of three phases which give zero potential difference; hence, if we can divide a more complex system into parts, some of which are those with zero potential difference, we can obtain general conclusions of great importance.

In order to determine which systems have zero potential difference, use can be made of the second law of thermodynamics. A convenient way of stating this law is in the form that work cannot be done continuously by using up the heat energy of the coldest part of the system.

We are concerned in our discussions with systems which are completely at one temperature, and so the second law of thermodynamics tells us that we cannot continuously do work if the only source of energy is heat energy.

In a system of three conductors, if the potential difference is not zero, then a current could flow round a complete circuit, and in doing so would perform work against any resistances that may be present. Thus, if a system of three conductors has a non-zero potential difference, there must be some source of energy, and this cannot, according to the second law of thermodynamics, be heat energy. If the passage of a charge right round the circuit leaves the chemical composition of each phase the same, then there can be no chemical energy to provide the work done; and, since the system remains exactly the same, there is no "running down," and so work is done continuously. If, in addition, there is no external supply of energy, then the only possible source of energy to supply the work would be heat energy, and this is precluded by the second law of thermodynamics. Thus we see that any system of three conductors, in which the passage of a current leaves the composition of the phases exactly the same, must be a system of zero potential difference, unless there is some external source of energy.

The first example that we can consider of a system with zero potential difference is that of a system of three metals. Here a charge passing round the circuit involves just the passage of electrons round, and hence there is no chemical change.

If we have a system of phases 1, 2, 3, 4 . . . n, 1, the total potential difference, $L_{1234\ldots n1}$ can be divided up into $L_{124\ldots n1}$ and L_{2342}. Now, let us suppose that the phases 2, 3, and 4 are all metals, so that L_{2342} is zero. We then see that the potential differences $L_{1234\ldots n1}$ and $L_{124\ldots n1}$ are equal. This gives the important result that: *the insertion of a metal between two other metals has no effect on the potential difference of a complete circuit*, the well-known "Volta's law," or the "law of intermediate metals" in thermoelectricity. This result explains why it is possible in, for example, a potentiometer circuit, to use any metal for the potentiometer wire, irrespective of the metals of the cells used.

If the phase 3 in the above discussion is not a metal, but a surface layer on a metal, the whole argument is unaltered, provided the phase behaves as a metal, *i. e.* provided the surface layer conducts by means of electrons. So we deduce that: *surface layers at a metal-metal junction have no effect on the potential difference of a complete circuit.* This explains why, again in a potentiometer circuit, there is no need to scrape off surface layers, provided that there is an electrical contact. In all the cases where measurements show that surface layers do have an effect on the potential difference it can be seen that it is not an effect of surface layers at metal-metal junctions, but surface layers at metal-dielectric junctions.

If, say, the phase 4 is not a metal but an electrolyte, then L_{2342} is the potential difference of a system of two metals and an electrolyte. In general, this is not zero, for there will be some chemical change when a current passes round the circuit, since the ions which enter the electrolyte will not usually be the same as those which leave it. Thus we shall not have the conditions which give rise to a zero potential difference.

But, if the phase 4 is an electrolyte and the phase 3 is the surface layer on the metal 2, then in the circuit of 2, 3, and 4 the transfer of a charge round the circuit may involve the transfer of one ion of metal 2 from the layer 3 into the electrolyte and the transfer of an ion from the electrolyte into the pure metal phase. If the surface layer remains the same, then there is no chemical change, and hence we have the conditions for zero potential difference, and using this result as above : *A surface layer on the metal at a metal-electrolyte junction may have no effect on the potential difference of a complete circuit.* This explains why consistent results are obtained for the potential differences of chemical cells without the elaborate precautions that are required as regards the removal of surface layers for consistency in other cases, *e. g.* the Volta effect. The above result also explains why, in the Daniell cell, electrodes of amalgamated zinc in place of the pure metal do not alter the potential difference, for, if we consider the amalgamated zone to be a surface layer on the pure metal, the above result can be used. The qualification "may have" is required, since there will clearly be some surface layers, *e. g.* of another metal, which do effect the potential difference.

A case that might, at first sight, seem to be in the category of systems of zero potential difference is that of two metals and an ionized gas, for here it is clear that there is no chemical change when the current passes. But the passage of the current discharges the ions at the metal surfaces, and so, for a continuous current, fresh ions must be supplied by some external ionizing agency, which takes the place of chemical change in providing the energy of the current.

In this connection, an interesting point has been raised by Villey (6) ; if we have a system of two metals and a gas at a fairly high temperature, then ionization of the gas may occur by thermal agitation alone, and this will give rise to a permanent current with no external source of energy for the ionization. If the current does work, then the energy for this comes from the heat energy of the gas, for this provides the energy of thermal agitation. This seems to be a contradiction of the second law of thermodynamics, for the energy of the current is supplied by the heat of the coldest part of the surroundings. The answer to the problem is that, owing to the removal of energy by the ionized molecules, the gas temperature is continually falling, and so the process is not continuous but gradually comes to an end.

In cases where we have a dielectric present, we cannot have the conditions for zero potential difference, since no current can flow round the circuit. Thus, for example, the system of a metal, its surface layer and a dielectric, has a definite potential difference, and so $L_{1234\ldots n1}$ and $L_{124\ldots,n1}$ are not the same. So: *surface layers at a dielectric boundary do affect the potential difference of a complete circuit.* There is thus a very important distinction between the effects of surface layers at conductor-conductor boundaries and at dielectric boundaries, and it is the neglect of this distinction which has led to such confusion as that of Hume-Rothery, quoted in the introduction.

9. *Potential Differences at a Boundary.*

In discussing potential differences at a boundary, we consider the passage of a charge from a point P to a point Q ; the passage of the charge is not to be considered as part of a continuous current, but as a " virtual " current across the boundary where there may be conditions of equilibrium, and the term " current across the boundary " must be taken to refer to this " virtual " current. In the complex cases where there is not true equilibrium, but merely electrical equilibrium, virtual currents in the two directions across the boundary will not involve just a reversal of a process, but will be processes carrying different substances across the boundary ; therefore, in obtaining potential differences in such cases, it is necessary to consider virtual currents in both directions and to use an average of these for obtaining " composite " potential differences at the boundary.

The process of carrying a charge across a boundary between phases 1 and 2 must start with the consideration of the carrier or carriers of charge (ions or electrons) in phase 1 ; consideration is then transferred to the carriers across the actual boundary, when these are not the same as those in phase 1 ; then we have to deal with the transfer of these carriers across the actual boundary ; and, finally, we consider the carriers in phase 2. Thus it is possible to divide up the whole process into a number of parts. It is convenient to use the term " potential difference " for each of these, extending the meaning to include the process of transferring consideration from one carrier to another in the same phase ; if we consider the change of thermodynamic potential in removing one carrier from the phase and replacing it by the other, we see that this is a " potential difference."

At any boundary there are a number of quantities which can be termed " potential differences," and it is necessary to be very careful in distinguishing these. For the purpose of defining the different potential differences, it is convenient to consider a boundary across which the current is carried by a single charged carrier, while in each phase the current is carried by a single carrier, of the same sign of charge as the carrier across the boundary, but of different nature. It is not a difficult matter to extend these definitions to cases where the carrier is of different sign or where there is more than one carrier. Carriers of charge in metals are electrons, and, therefore, it is more convenient to consider carriers charged negatively rather than positively.

As notation, we can use K_1 in connexion with the carrier of charge in phase 1 and ${}_2K_1$ in connexion with the properties in phase 1 of the carrier across the 1–2 boundary, ${}_1K_2$ being used for the properties of the same carrier in phase 2. A_{12} etc. represent potential differences across the boundary. With such notation it is seldom necessary to consider the individual carriers of charge.

If we consider, in any change, the effect of moving one coulomb of charge, then the potential differences must be measured in volts. If it is necessary to consider single electrons or ions, then potentials will be measured in electron-volts, and we can still refer to potential differences as a number of volts.

If the potential difference K_{12} is positive, then the usual sign convention is that the phase 1 carries a positive charge, so that if the measured potential difference $L_{123\ldots n1}$ is positive, then a positive current flows round the circuit in the direction 1, 2, 3 . . .n1. Since a phase becomes positively charged when a negative particle is removed from it, we adopt the convention that the removal of a negative particle is to be considered to correspond to a positive potential difference, so that the removal of a positive particle is a negative potential difference. It will be found that these conventions lead to results consistent with the conventions used by most European and many American workers in the subject.

10. *The Removal Potential Differences, D_1, ${}_2D_1$ etc.*

The change of the thermodynamic potential of the system due to the removal of an ion X^- from the phase 1 is defined as $D_{X^-,1}$, and is clearly the chemical potential μ_{X^-} of the constituent X^- in the phase 1. If X^- is the carrier of charge when a current flows in the phase 1, then $D_{X^-,1}$ can be written as D_1. But if X^- is the carrier across the boundary between phases 1 and 2, then $D_{X^-,1}$ is written ${}_2D_1$. It is clear that the only constituents for which a "D" is of interest will be those which carry the charge in the phase 1 or across the boundary from phase 1 to phase 2 or some other phase in contact with phase 1; the notation D_1, ${}_2D_1$ etc. is thus sufficient.

If the carrier across the boundary from phase 1 to phase 2 be a positively charged ion W^+, then we see that ${}_2D_1$ is not μ_{W^+}, but $-\mu_{W^+}$, owing to the sign convention of the previous section.

If more than one carrier takes part in carrying the current in the phase 1 or across the boundary, we have to take an average of the effects of the different carriers in order to obtain a "composite removal potential difference." If the current is carried by X^- and Y^+ ions, of mobilities v_- and v_+ respectively, then

$$D_1 = \frac{v_-\mu_{X^-} - v_+\mu_{Y^+}}{v_- + v_+}.$$

We can see that there must be a negative sign if we consider the special cases when v_- and v_+ are very small, so the current is carried almost entirely by ions of one sign; or we can consider the special case where $\mu_{X^-} = \mu_{Y^+}$ and $v_- = v_+$, and we see that the removal of 1 coulomb of negative charge must be accomplished by the removal of $\frac{1}{2}$ coulomb of X^- and the arrival of $\frac{1}{2}$ coulomb of Y^+; this must occur without changing the thermodynamic potential of the system, in agreement with the above formula.

We have not yet considered exactly what we mean by "removal" of an ion, whether it is removal to just outside the phase or to infinity; these will differ if the phase is charged, for there will be an electrostatic potential difference from the outside of the phase to infinity. It is convenient to use D_1 and ${}_2D_1$ for removal to just outside the phase, and we can define "just outside" by the specification that D_1 and ${}_2D_1$ are to retain the values they had for a neutral phase even when the phase becomes charged. If we wish to consider the removal of an ion to infinity, we can use the notation $\overline{D}_1$ and ${}_2\overline{D}_1$.

We shall leave until later the question of the relation between D_1 and the "photoelectric work function" in the case of metals where the carrier in the metal is an electron.

11. *The Carrier Change Potential Difference, ${}_2C_1$.*

If Z^- is the carrier of charge in phase 1 and X^- the carrier of charge across the boundary from phase 1 to phase 2, then, when we transfer consideration from Z^- to X^- we must remove one ion of Z^- from the phase and bring back one ion of X^-; these two processes involve changes of potential D_1 and ${}_2D_1$ and we can define the carrier change potential difference ${}_2C_1$ by

$${}_2C_1 = D_1 - {}_2D_1.$$

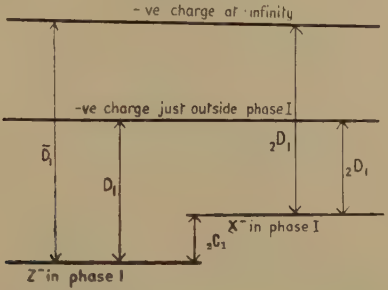
The notation ${}_2C_1$ is unambiguous, for it concerns the change from the carrier in the phase 1 to the carrier across the boundary from phase 1 to phase 2.

Where there are two or more carriers, D_1 or ${}_2D_1$ must be defined, as above, as a "composite" or "average" removal potential difference.

It may be noted that ${}_2C_1 = \overline{D}_1 - {}_2\overline{D}_1$, since the change of potential in taking Z^- from just outside phase 1 to infinity is the same as the change of potential in bringing X^- back from infinity to the same point.

The relation is illustrated in fig. 1.

Fig. 1.


- ve charge at infinity
-ve charge just outside phase I
$\bar{D}_1$
D_1
${}_2D_1$
${}_2D_1$
X^- in phase I
${}_2C_1$
Z^- in phase I

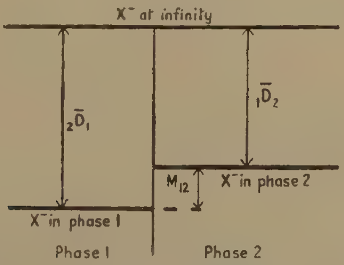
The removal and carrier change potential differences.

12. *The Transfer Potential Difference, M_{12}.*

When the carrier across the boundary moves across, there is a change of potential of the whole system denoted by M_{12}. If there is equilibrium at the boundary, it is clear that M_{12} will be zero, but, since we have to consider cases where the boundary is that between a conductor and a dielectric, it is clear that we have to retain the quantity M_{12} in a general discussion.

We see that M_{12} is given by the difference between ${}_2\overline{D}_1$ and ${}_1\overline{D}_2$, for we can go from one phase to the other by removing the carrier to infinity and then returning it. M_{12} is thus illustrated by fig. 2.

Fig. 2.

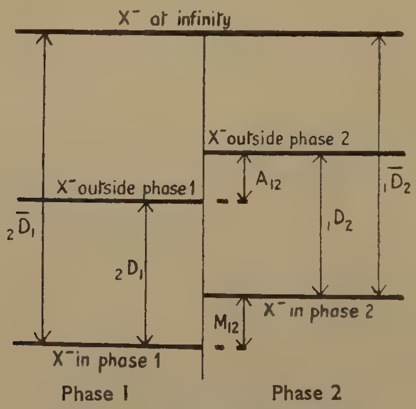

X^- at infinity
${}_2\bar{D}_1$
${}_1\bar{D}_2$
M_{12}
X^- in phase 2
X^- in phase 1
Phase 1
Phase 2

The transfer potential difference, M_{12}.

13. *The External Potential Difference, A_{12}.*

If we consider the difference of potential as between a particle X^- just outside phase 1 and just outside phase 2, we get the external potential difference, A_{12}. The relations between this and the other potential differences are seen by inserting in fig. 2 levels corresponding to X^- outside phases 1 and 2, giving fig. 3.

Fig. 3.



The external potential difference, A_{12}.

We see that $A_{12} = {}_1D_2 - {}_2D_1 + M_{12}$.

But $M_{12} = {}_2\bar{D}_1 = {}_1\bar{D}_2$.

So $A_{12} = {}_2\bar{D}_1 - {}_2D_1 - ({}_1\bar{D}_2 - {}_1D_2)$.

If there is equilibrium at the boundary, so that M_{12} is zero, we get

$$A_{12} = {}_1D_2 - {}_2D_1.$$

A_{12} is the potential difference across the actual phase boundary, and can be represented by true electrostatic lines of force across the boundary. If there is any relative motion of the two phases, or slip along the boundary, it must be the lines of force of the potential difference A_{12} which have relative motion of the two ends, and hence it is A_{12} which is related to the "electrokinetic potential" defined by Freundlich (7). In many cases where there is an electrokinetic potential, there will be surface layers giving a number of different phase boundaries, at each of which there is an external potential difference, but, usually, at only one of which is there any relative motion, and hence an electrokinetic potential. Freundlich and others have found that there is little connexion between the electrokinetic potential and the so-called "electrode potential"; even apart from the fact that, as will be discussed in the part on the Voltaic cell, the sum of the external potential differences at the boundaries may not be the electrode potential, there is no reason to expect that the one external potential difference at the boundary of relative motion should have much connexion with the sum of the potential differences for all the boundaries concerned.

14. *The Internal Potential Difference, B_{12}.*

If we consider the whole process of the passage of a current from one phase to another, we can define a potential difference B_{12}. This can be considered as built up by the potential difference, ${}_2C_1$, of the change from the carrier of charge in phase 1 to that across the boundary from phase 1 to phase 2, followed by the potential difference, ${}_1M_{12}$, of the transfer of the charge from one phase to the other (zero if there is equilibrium), and finally the potential difference, ${}_1C_2$, of the charge from the carrier of charge across the boundary to that in phase 2. Thus

$$B_{12}={}_2C_1+M_{12}-{}_1C_2.$$

But, as discussed previously,

$$ {}_2C_1=D_1-{}_2D_1 $$

and

$$ {}_1C_2=D_2-{}_1D_2. $$

So

$$B_{12}=D_1-D_2+M_{12}-{}_2D_1+{}_1D_2.$$

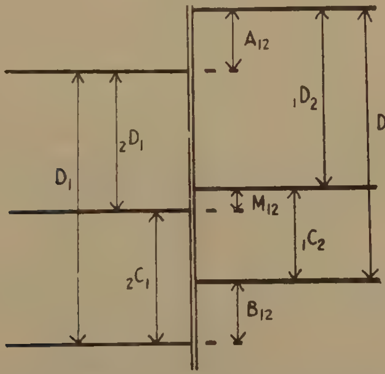
But

$$A_{12}={}_1D_2-{}_2D_1+M_{12},$$

and so we have the alternative result,

$$B_{12}=D_1-D_2+A_{12}.$$

Fig. 4.



Relationship of potential differences.

This can be understood if we consider the transfer of charge across the boundary to be accomplished by removing a charge from phase 1, potential difference D_1, taking a charge (of any nature) across the boundary outside the phases, potential difference A_{12}, and then returning a charge to phase 2, potential difference $-D_2$.

We see that the two interphase potential differences, A_{12} and B_{12}, will only be the same if $D_1=D_2$, that is, if the change of potential is the same for the removal of a carrier from either phase, a case which would only occur by chance.

Fig. 4 shows the relationship of the various potential differences.

15. *Analogy for the Potential Differences.*

It may assist in the understanding of the potential differences if we can have some sort of analogy for them.

A simple analogy can be found in the consideration of an underground railway system and, in particular, of two stations of this with a connecting line, there being another line to each station. Each line is to be considered to be horizontal, but all are at different levels. The depth of a platform below the surface represents a removal potential difference such as D_1; the difference in level of the two lines at one station represents a change of carrier potential difference such as ${}_2C_1$; the difference in level of the surfaces at the stations represents the external potential difference A_{12}, while the internal potential difference B_{12} is represented by the difference in level of the two lines that do not connect the two stations. If the connecting line is not horizontal, its difference of level at the two ends represents M_{12}. It is clear that the meaning of the term "difference of level" between the two stations depends on the viewpoint. To the traveller using only the intermediate line there is the difference in level, M_{12}, but if he uses the intermediate line to connect up journeys on the other two lines, he will consider the difference of level to correspond to B_{12}. On the other hand, the surveyor or pedestrian would be much more interested in the analogue of the potential difference A_{12}.

16. *The Total Potential Differences of a System.*

If we consider the virtual passage of a charge right round a system of three junctions (and, as we have seen, a more complex system can always be resolved into systems of three junctions each), we can get the measured potential difference of the whole system. If we consider the internal potential differences, B_{12}, B_{23}, and B_{31}, we can see that the sum of these must give the measured potential difference, for there can be no other change of potential in going right round the circuit, so that

$$L_{1231}=B_{12}+B_{23}+B_{31}.$$

But

$$B_{12}=D_1-D_2+A_{12},$$

and so we see that

$$L_{1231}=A_{12}+A_{23}+A_{31}.$$

the terms D_1, D_2, and D_3 cancelling.

We thus reach the important conclusion that: *the total potential difference of a complete circuit can be divided either into the internal potential differences or into the external potential differences*;

or:

the sum of the internal potential differences of a complete circuit is equal to the sum of the external potential differences.

17. *Conclusion.*

We have seen that there are two distinct ways of defining interphase potential differences, and that these lead to the same result—that the sum of the interphase potential differences is equal to the measured potential difference of the complete system. Since it is only possible to measure potential differences when there are three or more phases in the system, it might seem that there is little point in discussing interphase potential differences ; but, in fact, in some cases the interphase potential differences can be calculated and in other cases they can be deduced from measurements, thus leading to conclusions which, as has already been pointed out, may prove false if care is not taken to have a consistent meaning for the " interphase potential difference."

References.

(1) Kelvin, Phil. Mag. xlvi. p. 82 (1898).
(2) Lodge, B.A. Report (1885) ; Phil. Mag. v. p. 19 (1885) ; xlv. pp. 351, 454 (1900).
(3) Hume-Rothery, ' The Metallic State,' p. 148.
(4) Lange and Nagel, *Zeitsch. Elektrochem.* xlii. p. 50 (1936).
(5) Guggenheim, ' Modern Thermodynamics by the Methods of Willard Gibbs.
(6) Villey, *Journ. de Phys.* iv. p. 10 (1933).
(7) Freundlich, ' New Conceptions in Colloidal Chemistry,' p. 18.

XLII. *Contact Potentials.*—Part II. *The Volta Effect.*

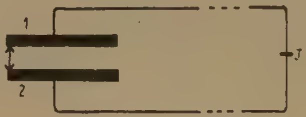
By J. Alan Chalmers, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham *.

[Received February 4, 1942.]

1. *Introduction.*

In Part I. the general principles of the theory of contact potentials have been discussed, and definitions have been given of two interphase potential differences, the external potential difference, denoted by A_{12}, and the internal potential difference, denoted by B_{12}. In this part these principles are used to discuss the Volta effect and to elucidate some of the difficulties that occur in the previous treatments of the subject.

Fig. 1.



The Volta effect.

The Volta effect comprises the measurement of the potential difference of a system consisting of two metals and a dielectric, usually air or a vacuum. It is probable that seldom, if ever, is it possible to obtain a system with just the three phases, for there will be surface layers on each metal unless very special precautions are taken. As has been discussed in Part I., surface layers at the metal-metal contact have no effect, but at the metal-dielectric contacts they have a great influence.

2. *Experimental Methods.*

In the simplest arrangement for demonstrating the Volta effect, as shown in fig. 1, we have two plates of the metals, joined by wires of negligible capacity to a distant junction J ; the potential difference of the Volta effect is the electrical potential difference between two points just outside the two plates ; if this is V, then the charge on each plate is cV, where c is the capacity of the system ; if, therefore, the capacity c is altered, for example, by altering the distance apart of the plates, a charge

---

* Communicated by the Author.

will flow round the circuit. A measuring instrument inserted between the plate 1 and the junction J can measure this charge, and if the change of capacity is known, then the potential difference V is obtained. This simple method was used by Kelvin (1), who improved on the method by inserting in the circuit also a variable source of potential (cell and potentiometer) and adjusting this so that no charge passes, in which case the potential applied must be just equal and opposite to the potential difference of the Volta effect.

If there is a gas between the plates 1 and 2, and this is ionized, then a current will flow across the gas unless the Volta potential difference is opposed by an equal and opposite potential difference applied as above. Although the ionization of the gas really converts the system into one of two metals and an "electrolyte," the results are the same as obtained by the electrostatic method; the reason for this agreement will be discussed later.

The passage of thermionic currents between a filament and an anode depends on the potential difference between the two, and this potential difference must include the Volta potential difference between the two metals as well as the applied potential difference. This has been used to determine the Volta potential difference *in vacuo*.

As will be shown later, the Volta potential difference between two metals is related to the "photoelectric work function," and Millikan (2) has measured the value of this for various metals.

3. *General Experimental Results.*

The results for the Volta potential difference by all methods show that, for any given pair of surfaces, there is a definite potential difference which is independent of the sizes of the surfaces and of their distances apart, but which does depend on the actual nature of the surface layers, and is altered by anything which alters the nature of the surfaces.

If the surfaces are left undisturbed, there is found to be no effect on reducing the pressure or on replacing the air by other gases which have no chemical effect on the metals; for example, Bottomley (3) reduced the pressure from atmospheric to 1/523 mm. of mercury without an alteration in the Volta effect of as much as 1 per cent.

On the other hand, when the surfaces at the metal-dielectric contact are disturbed, the Volta potential difference is much affected. For example, Erskine-Murray (4) altered the potential difference by 50 per cent. by polishing and scratching, and Spiers (5) found very large effects (nearly 1 volt change) by heating iron in hydrogen.

Rose (6) found a potential difference of about 0·4 volts between the 100 and 111 faces of single copper crystals, and this may account for many of the effects of polishing, etc.

4. *The Location of the Potential Difference.*

The potential difference of the Volta effect is the potential difference of a system with three junctions, two of metal-dielectric and one of metal-metal. Attempts have therefore been made to locate the potential difference either at the two metal-dielectric junctions or at the metal-metal junction.

In view of the discussion of Part I. as to the different potential differences that can be defined at a boundary, it follows that the location of the "Volta potential difference" at one or more boundaries depends on how this is defined in terms of the potential differences. As we have seen in Part I., the total potential difference of the system can be divided up either as internal or as external potential differences.

Many arguments have been brought forward to support each of the theories of the location of the potential difference, and the inconclusive result of the discussions has led to many ambiguities. In order to clear up this question, it seems desirable to give the two theories in the forms in which they were originally discussed and the arguments on both sides; then we shall see how the two theories fit naturally into place according to the precise definition of "potential difference" that is adopted.

5. *The Contact Theory.*

Volta himself considered the potential difference to originate at the metal-metal contact, and this view has been upheld by Kelvin (1) and others. According to this theory, the whole potential difference of the Volta effect is located at the metal-metal boundary; in order to account for the variations in the potential difference due to alterations in surface conditions, it is necessary to modify the theory to allow of part of the potential difference existing at the metal-surface layer boundary near the metal-dielectric boundary.

An important argument in favour of the contact theory is the fact that there is a fairly close correspondence between the potential differences of the Volta effect and the Voltaic cell for the same pair of metals; the common factor here is the metal-metal boundary, and it seems that this is a good reason for believing that the metal-metal boundary is the location of the Volta potential difference.

Since the Volta effect persists in the highest vacuum attainable, it seems certain that the potential difference cannot originate in chemical action at the metal-dielectric boundaries. But, as no continuous current flows in the Volta effect, there is no need to seek for a source of energy to move charges continuously.

Against the contact theory has been urged the fact that, if there is a large potential difference at the metal-metal boundary, then the corresponding amount of work must be done to transfer a charge from one metal to the other ; whereas, in fact, a charge can be transferred with only an amount of work (as shown by the heating), namely the Peltier effect, which is very much smaller in magnitude than that corresponding to the Volta potential difference. However, Kelvin (1) has shown thermodynamically that, if π_{12} is the heating effect and V_{12} the Volta potential difference, then $\pi_{12}=T\frac{dV_{12}}{dT}$ and π_{12} may be much smaller than V_{12}.

6. *The Chemical Theory.*

De la Rive (8) first brought forward the theory that the Volta potential difference originates at the metal-dielectric contacts, due to chemical action there. This idea was energetically supported by Lodge (9) and others.

In support of the chemical theory is the undoubted fact that the energy of the Voltaic cell is known to come mainly from the chemical action which occurs at the electrodes ; this leads to the conclusion that the potential difference also occurs at the electrodes, and so that there is no potential difference at the metal-metal junction. If the metal-metal junction in the Voltaic cell provides no part of the potential difference, neither can it do so in the Volta effect, so that the potential difference in this case must originate at the metal-dielectric junctions. In order to account for these potential differences at metal-dielectric junctions, there must be assumed to be some sort of chemical action at these junctions to correspond to the chemical action at the electrodes ; and, since the potential differences of the Volta effect and Voltaic cell are similar in magnitude, the chemical processes at work in the two cases must be somewhat similar. Lodge suggested that in both cases the process is one of oxidation of the metal by air or water ; de Broglie (10) thought it was the interaction of the metal with water. The discovery that the Volta effect persists in a high vacuum was a serious blow to the chemical theory, though it did not answer some of the arguments in favour of this theory.

In favour of the chemical theory is the argument brought forward in regard to the Peltier heat ; even with Kelvin's proof that the Peltier effect would not be expected to be of the same magnitude as the potential difference at the metal-metal boundary, there is still no satisfactory answer to the question as to what becomes of the energy that must be liberated if a charge moves through a potential difference such as is understood, on the contact theory, to exist at the metal-metal boundary.

Another very convincing argument in favour of the chemical theory as against the contact theory lay in the effects of surface layers. If the surface layers affect the Volta potential difference, we might well expect that it is those surface layers at the location of the origin of the potential difference which, alone, are likely to affect the Volta potential difference. Now, unless chemical changes actually occur there, surface layers at a metal-metal boundary have no effect, whereas surface layers at the metal-dielectric boundary do have an effect. Hence there seems good reason for locating the origin of the potential difference at the metal-dielectric boundaries, as postulated by the chemical theory.

7. *Discussion of the Two Theories.*

It can be seen that many of the arguments in favour of each theory are, in fact, arguments against the alternative theory, rather than any definite evidence in favour of one theory.

The fact that the Volta potential difference persists in a vacuum is a definite argument against this potential difference being due to chemical action at the metal-vacuum junction, but it is not necessarily evidence that the potential difference exists only at the metal-metal junction.

Again, the argument with regard to the Peltier heat is an argument against the potential difference existing at the metal-metal junction, but not any evidence that it does exist at the metal-dielectric junctions.

The argument with regard to the surface layers ceases to have any importance when it is realized that the results obtained experimentally can be deduced, as in Part I., without any discussion at all as to how the potential difference is to be located, but merely from the nature of the conduction in different media.

It will be seen that the conflicting points of view can be reconciled if care is taken in defining what is meant by a " potential difference " when located at a single boundary. It will be seen that the potential difference is located at the metal-metal boundary if by potential difference we mean what has been defined in Part I. as the external potential difference, while if the potential difference is the internal potential difference then it is at the metal-dielectric boundaries that we find potential differences to exist. And it will be seen that the arguments against each theory will cease to hold with the exact definitions in which the theory is true.

In order to obtain a correct interpretation of the Volta effect, it is necessary to consider each of the phase boundaries concerned and then to see how the whole potential difference is built up of the potential differences at the different boundaries. Since a surface layer behaves like a metal, as far as conduction is concerned, the discussion of a metal-surface layer boundary is similar to that of a metal-metal boundary, while the discussion of a surface layer- dielectric boundary is like that of a metal-dielectric boundary. So the effects of surface layers can be introduced into the discussion very simply.

The Volta effect is most simply discussed for two metals and a vacuum, leaving until later the alterations to be made due to the replacement of the vacuum by a dielectric with dielectric constant differing from unity.

8. *The Metal-Vacuum Boundary.*

In a metal, the current is carried by electrons, and it is therefore convenient to choose electrons when we imagine charges to be moved in the vacuum, in discussing potential differences across the vacuum. We can now define the various potential differences which occur at the boundary. In the metal (denoted as phase 1) there is a removal potential difference D_1 which is almost the same as the photoelectric or thermionic work function (the difference will be discussed later); since the carriers across the boundary into the vacuum (phase 3) are to be considered electrons, which also carry the current in the metal, ${}_3D_1=D_1$, and hence the change of carrier potential difference ${}_3C_1=0$. In the vacuum, the removal potential difference D_3 is zero since we cannot discuss the removal of an electron from a vacuum to any other place "outside the vacuum"; and ${}_1C_3$ must also be zero.

In order to obtain the interphase potential differences, it is necessary to know one of them, and in this case it is easy to see that the external potential difference A_{13} must be zero, because this is concerned with moving a charge from outside phase 1 to outside phase 3, along the lines of electric force which must end on charges in either phase; since there can be no charges situated in the vacuum, there can be no such lines of force and hence no external potential difference A_{12}. Or we can see that the change of potential in moving an electron from the metal, namely D_1, must be the same as moving it from the metal into the vacuum, namely, the transfer potential difference M_{13}.

Using the results obtained in Part I., we see that the internal potential difference B_{13} must be equal to D_1, the other terms A_{13} and D_3 being zero.

In the same way, for the second metal-vacuum boundary,

$$A_{23}=0$$

$$M_{23}=B_{23}=D_2.$$

9. *The Metal-Metal Boundary.*

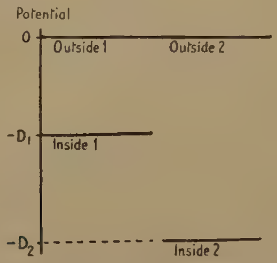
Since electrons carry the current in both metals and across the boundary, there is no change of carrier at the boundary, and so ${}_1C_2$ and ${}_2C_1$ are both zero. At a metal-metal boundary, the electrons can pass freely across the boundary (the "potential hill" being overcome by "tunnelling" as in the theory of radioactive disintegration), and hence there will be equilibrium, and so M_{12} will be zero.

Since ${}_2C_1$ and ${}_1C_2$ are zero, ${}_2D_1=D_1$ and ${}_1D_2=D_2$, so that $A_{12}=D_2-D_1$.

If we have two metals 1 and 2, each electrically neutral and insulated, we can represent potentials as in fig. 2.

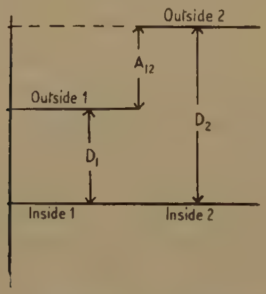
If we now imagine the two metals to be joined, electrons will tend to flow from the metal 1 to the metal 2, and it is clear that a state of equilibrium will be reached when the potentials of the electrons inside the two metals will be equal; this equalization of potential is achieved by the charging of the two metals, 1 positively and 2 negatively, and can be represented by lowering the levels for metal 1 and raising those for metal 2, giving fig. 3.

Fig. 2.



Potentials of electrons in neutral metals.

Fig. 3.



Potentials of electrons in metals in contact.

These diagrams show clearly how the external potential difference, A_{12}, must be equal to D_2-D_1.

10. *The Total Potential Difference of the System.*

In Part I. we have seen that the total potential difference of a system can be obtained either as the sum of the external potential differences or the sum of the internal potential differences. Thus the measured Volta potential difference V_{12} is given by:

$$V_{12}=A_{12}+A_{23}+A_{31},$$

or

$$V_{12}=B_{12}+B_{23}+B_{31}.$$

We have found :

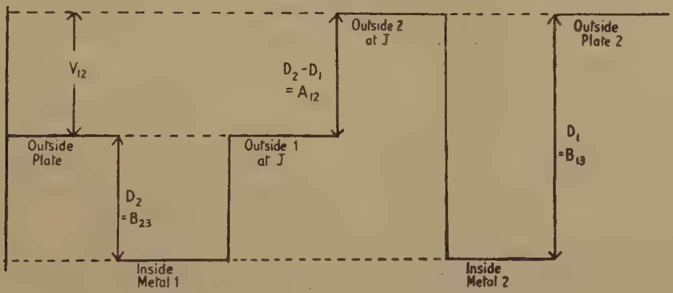
$$A_{13}=A_{23}=0,\ A_{12}=D_2-D_1,$$
$$B_{13}=D_1,\ B_{23}=D_{21}B_{12}=0.$$

Thus, from either method, as illustrated in fig. 4,

$$V_{12}=D_2-D_1.$$

If the total potential difference of the system be taken as the sum of the external potential differences, then there is no potential difference at the metal-vacuum boundaries, the whole being concentrated at the metal-metal boundary. But if the total potential difference be taken as the sum of the internal potential differences, then there will be potential differences D_1 and D_2 at the two metal-vacuum boundaries and none at the metal-metal boundary. These two methods of dividing up the Volta potential difference correspond to the two theories of the Volta effect.

Fig. 4.



The potential differences of the system.

11. *The Peltier Effect.*

Since the Peltier effect has been considered in discussing the two theories, it is necessary to see how this fits into the picture we have given.

When there is equilibrium at a metal-metal junction, there is no change in thermodynamic potential when an electron crosses the boundary, but this does not necessarily mean that there are no heat effects. If there is a change in the entropy per electron as an electron passes across, then this will mean the liberation or absorption of heat, *i. e.*, a Peltier effect.

If G_1, G_2 are the thermodynamic potentials per electron in the neutral metals, and $\overline{G}_1$, $\overline{G}_2$ when the metals are charged, then $\overline{G}_1=\overline{G}_2$ and G_1, G_2 are $-D_1$, $-D_2$ respectively.

Now $G=H-TS$, where H is the "total heat," T the temperature, and S the entropy, and $S=-\left(\frac{\partial G}{\partial T}\right)_{p,n}$.

So $G_1-G_2=H_1-H_2-T(S_1-S_2)=D_2-D_1=A_{12}=H_1-H_2-T(S_1-S_2)$.

And $T(S_1-S_2)=-\frac{\partial}{\partial T}(G_1-G_2)=-\frac{\partial}{\partial T}A_{12}=-\frac{\partial}{\partial T}V_{12}$.

So the Peltier heating, which is $T(S_2-S_1)$ is given by the rate of change of the total potential difference with the temperature, as deduced by Kelvin (7).

12. *The External Potential Differences and the Contact Theory.*

Since the contact theory ascribes the whole potential difference of the Volta effect to the metal-metal contact, it is clear that we can obtain agreement with this theory if we divide up the whole potential difference of the Volta effect into parts corresponding to the external potential differences.

The argument against the contact theory, in connexion with the fact that in the analogous Voltaic cell there is chemical action at the electrodes, must be left to Part III., when the Voltaic cell is discussed, and it will be seen that this no longer holds.

The argument against the contact theory, in regard to the smallness of the Peltier effect, does not hold, since the passage of an electron across the boundary does not involve any total change of potential, in spite of there being an external potential difference.

The other argument against the contact theory, in regard to the fact that surface layers have no effect at the metal-metal boundary, has already been discussed and will be further considered later.

All the arguments in favour of the contact theory are in agreement with the interpretation of the potential difference in terms of external potential differences.

13. *The Internal Potential Differences and the Chemical Theory.*

The chemical theory locates the potential difference at the two metal-vacuum junctions, and we can obtain this result if the potential difference is divided up according to the internal potential differences, D_1 and D_2, at the two metal-vacuum boundaries, with none at the metal-metal boundary.

Although, in this way of regarding the matter, the metal-metal junction does not contribute to the potential difference, the fact that the Volta effect and Voltaic cell are similar in magnitude for the same metals can still be understood, as we shall see in the discussion of the Voltaic cell in Part III.

The potential differences D_1 and D_2 do not, in themselves, represent chemical effects, and hence it is not correct to call the theory a " chemical " theory. But D_1 and D_2 persist in a vacuum, and hence the objection is no longer valid that there can be no potential difference due to chemical action at a metal-vacuum boundary.

The interpretation of the Volta effect in terms of internal potential differences gives agreement with the requirements of the chemical theory, modified so as not to have true chemical effects but, rather, photoelectric effects, and the objections to the chemical theory can be overcome.

14. *The Metal-Metal Potential Difference.*

According to which theory is made use of, there is either a metal-metal potential difference of magnitude $V_{12}=A_{12}$, the whole of the Volta potential difference, or else a metal-metal potential difference which is zero. Therefore it seems very desirable not to use the term " metal-metal potential difference " without describing the potential difference as " internal " or " external."

Confusion has often arisen because the arguments of the chemical theory have been accepted as proving the absence of the metal-metal potential difference, and then this result has been made use of under conditions in which it is not applicable. The arguments show that B_{12} is zero, and this can only be used in other discussions in which the internal potential differences are used ; but often it is, in fact, the external potential differences which are used in the discussions, and it is sometimes not realized that A_{12} is not zero. We shall see examples of this when dealing with the Voltaic cell, which includes a metal-metal boundary.

It does not seem profitable to discuss whether the contact theory or the chemical theory provides a better interpretation of the Volta effect, that is to say whether it is preferable to divide up the total potential difference of the system into external or internal potential differences ; for different purposes, either may be the more suitable method of division. All that is essential is to prevent the two theories from becoming mixed.

15. *Surface Layers.*

We have already seen in Part I. that surface layers at a metal-metal junction can have no effect on the total potential difference of a system. If we denote the surface layers on the metals 1 and 2 by 1′ and 2′, we can interpret this result as either

$$A_{12}=A_{11'}+A_{1'2'}+A_{2'2},$$

or

$$B_{12}=B_{11'}+B_{1'2'}+B_{2'2}.$$

But when we deal with surface layers at a metal-vacuum boundary, a similar result does not hold and we can say that

$$A_{13} \neq A_{11'} + A_{1'3}$$

and

$$B_{13} \neq B_{11'} + B_{1'3}.$$

If we consider the surface layers as metals with removal potential differences $D_{1'}$ and $D_{2'}$, differing from D_1 and D_2, we can interpret these results.

Just as A_{13} is zero, so is $A_{1'3}$, but $A_{11'} = D_{1'} - D_1$; and $A_{1'2'}$ and $A_{2'2}$ are $D_{2'} - D_{1'}$ and $D_2 - D_{2'}$ respectively, giving the above results.

And just as B_{12} is zero, so are $B_{11'}$, $B_{1'2'}$ and $B_{2'2}$, but $B_{13} = D_1$ and $B_{1'3} = D_{1'}$.

Using either external or internal potential differences for a system of vacuum, surface layer, metal, two surface layers, metal, surface layer vacuum, we get

$$V = D_{2'} - D_{1'},$$

which depends entirely on the properties of the surface layers and not on the properties of the pure metals at all, a result which could also be deduced from Volta's law that an intermediate metal has no effect on the potential difference of a system.

In locating the potential difference of such a system, we must have external potential differences $A_{11'}$ and $A_{2'2}$ at the bases of the surface layers at the metal-vacuum boundaries, so that the potential difference can no longer be located entirely at a metal-metal junction.

16. *Dielectrics.*

When there is a dielectric present, rather than a vacuum, some of the assumptions made above may no longer be correct, and it is necessary to discuss what effect these may have upon the Volta effect

The energy of a charge $\pm e$ situated on a carrier of radius r in a medium of dielectric constant k is $\frac{1}{2}\frac{e^2}{kr}$, and hence there is a change of potential D_3 in removing an electron from the dielectric to a vacuum where its energy is $\frac{1}{2}\frac{e^2}{r}$. Thus $D_3 = \frac{1}{2}\frac{e^2}{r}\frac{(k-1)}{k}$. Since this occurs at both ends of a dielectric, which replaces air, it will cancel out in considering the whole potential difference of the Volta effect, unless, of course, the dielectric, by chemical effects, produces surface layers on the metals.

We see that the replacement of a vacuum by any dielectric produces no alteration in the Volta effect, so that the system metal/vacuum/dielectric must have zero potential difference. Similarly, a system of metal and two dielectrics has zero potential difference, and so it can be deduced also that a system of three dielectrics has zero potential difference.

If D_3 is now no longer zero, then the transfer potential difference M_{13} is now $D_1+A_{13}-D_3$ and, as there are no changes of carrier, $B_{13}=M_{13}$.

It is no longer correct to say that the external potential difference A_{13} is zero, for there will be some charge present on each edge of the dielectric and produced by the polarization of the dielectric. Simple dielectric theory shows that if the total potential drop across the dielectric is V_{12}, then the external potential differences at the two ends of the dielectric amount to a fraction $\frac{k-1}{k}$ of this. But, as we have seen, this makes no difference to the Volta effect, nor to any other experimental determination, and it remains possible to discuss the Volta effect with a vacuum and then to use the above results for the replacement of a vacuum by a dielectric. We may note that, if k becomes very large, the whole of the potential difference, V_{12}, becomes concentrated at the two ends of the dielectric.

17. *Ionized Gases.*

It is an experimental fact that the Volta potential difference, as measured by the ionization method, is the same as by other methods, and we now have to consider how the same measured potential difference can be obtained for what is actually a system of two metals and an "electrolyte," as for two metals and a dielectric.

For simplicity we will consider a gas which forms positive ions by the removal of one electron from a neutral molecule and negative ions by the addition of one electron to a neutral molecule. Let us suppose that the change of potential in removing an electron from a neutral molecule, leaving a positive ion and an electron, is J, and the change of potential in removing an electron from a negative ion, leaving a neutral molecule and an electron, is J′. Then the total change of potential in forming a pair of ions is $I=J-J'$.

At the boundary between the metal and the ionized gas, the carrier across the boundary must be an electron; thus there is no carrier change potential difference ${}_3C_1$ in the metal. But the carriers in the ionized gas must be gaseous ions, and so there is a ${}_1C_3$, and there will be two removal potential differences in the gas, D_3 for the ions and ${}_1D_3$ for electrons.

Let us define P_+ and P_- as the changes of potential in removing one positive and one negative ion respectively from the gas. Now let us consider the change of potential in removing one positive ion and one electron from the gas; if these are removed separately, the change of potential is $P_+ + {}_1D_3$, but we can reach the same result by allowing the two to recombine, potential change $-J$, removing the neutral molecule, potential change W, and then separating the positive ion and electron, potential change J. Thus $P_{+}+{}_{1}D_{3}=W$. Similarly, $P_{-}-{}_{1}D_{3}=W$.

Now, if we assume that the two ions have equal mobilities, we can define the "composite removal potential difference," D_3 (see Part I. section 10), where

$$D_3=\frac{P_{-}-P_{+}}{2}={}_{1}D_{3}.$$

Then we have $${}_{1}C_{3}=D_{3}-{}_{1}D_{3}=0.$$

Thus $B_{13}=M_{13}$ still, since both carrier change potential differences are zero. So all the results of sections 8, 9, and 10 still hold, and the potential difference of the whole system is the same as if the gas were not ionized.

18. *The Photoelectric Effect.*

When light of frequency ν falls on a metal plate, the energy of the electrons emitted by the photoelectric effect is $h\nu-Q$, where Q is a constant for the metal known as the "photoelectric work function." At first sight, this might be expected to be the same as the removal potential difference, D_1, for the removal of an electron from the metal; but this is not quite so. Applying the Thomson equation, $G=H-TS=H+T\frac{\partial G}{\partial T}$ or, for changes in the quantities used under conditions of constant temperature, $\Delta G=\Delta H-T\Delta S=\Delta H-\Delta Q$, the removal potential difference D_1 is ΔG; but the work required to be provided from outside is ΔH, the difference ΔQ coming from the source of heat which maintains the temperature constant, the argument being similar to that of the chemical cell; so Q, the photoelectric work function, is

$$\Delta H=\Delta G-T\frac{\partial \Delta G}{\partial T}=D_1-T\frac{\partial D_1}{\partial T}.$$

Sommerfeld (11), following Dushman (12), has used the thermodynamic potential to deduce the formula for thermionic emission; he finds a formula involving $\frac{G}{n}$ the thermodynamic potential per electron, which can be interpreted as the "thermionic work function" and is the same as we have called D_1. The formula is $i=aT^2e^{-b/kT}$, where $b=D_1$; but the experimental determination of the thermionic work function is obtained from the slope "s" of the curve relating $\log i-2\log T$ with $-\frac{1}{kT}$.

We see that $$s=b-T\frac{\partial b}{\partial T},$$

as obtained by Becker and Brattain (13).

We see that $s=\frac{\partial H}{\partial n}=Q$, which is the true photoelectric work function. Becker and Brattain have preferred to call $s=Q$ the "heat function" and D_1 the "photoelectric work function," but the distinction is merely one of nomenclature, and the difference between the two is usually small.

Millikan (2) has related the photoelectric effect to the Volta effect by measuring the potential U_{12}, which must be applied between two plates to stop electrons emitted from plate 1 by light of frequency ν, so that they are just unable to reach the plate 2. In addition to the applied potential U_{12} there is the potential difference V_{12} of the Volta effect, so that $U_{12}+V_{12}=h\nu-Q_1$. Replacing the plate of metal 1 by another of metal 3, and using the same light, Millikan measured U_{32} and found, to the limit of his experimental error, that U_{32} was always the same as U_{12}. Therefore $V_{12}-V_{32}=Q_3-Q_1$. But, using Volta's law, or the fact that a system of three metals has no total potential difference, $V_{12}-V_{32}=V_{13}$. We have already seen that $V_{13}=D_3-D_1$, so that Millikan's experiments show us that Q_1 and D_1 do not differ by as much as his experimental error, and we can see, by comparison with section 11, that the difference between D_3-D_1 and Q_3-Q_1 is the small Peltier effect π_{13}.

It would seem that measurements of the Q's would give an accurate method of determining the Volta effect, but there is one difficulty, namely, that the photoelectric work function of a metal depends on the particular crystal orientation that is used. If we use measurements of Q, *e. g.*, from the minimum energy required to remove an electron, *i. e.*, the threshold frequency, and carry these out for a polycrystalline metal, we are actually measuring, not the average Q_1 for all the crystals, but rather the Q_1 for that particular crystal orientation which gives the minimum Q_1. But if we consider other measurements of the Volta effect, we see that it is the average Q_1 or D_1 that is required, and so may not agree with the photoelectric measurements.

19. *Conclusion.*

We have seen that the phenomena of the Volta effect can be given a consistent interpretation in terms either of the external potential differences, agreeing with the contact theory, or of the internal potential differences, agreeing with the chemical theory ; each way of regarding the matter is correct provided the ideas of each are kept separate. We have seen, too, that the effect of replacing a vacuum by a dielectric does not affect the Volta potential difference, and that the ionization method gives the same result as other methods ; the relationship of the Volta effect with the Peltier effect and the photoelectric work function has also been discussed.

References.

(1) Kelvin, Phil. Mag. xlvi. p. 82 (1898).
(2) Millikan, Phys. Rev. vii. pp. 18, 355 (1916) ; xviii. p. 236 (1921).
(3) Bottomley, B. A. Report, lv. p. 901 (1886).
(4) Erskine-Murray, Proc. Roy. Soc. lxiii. p. 113 (1898).
(5) Spiers, Phil. Mag. xlix. p. 70 (1900).
(6) Rose, Phys. Rev. xliv. p. 585 (1933).
(7) Kelvin, Phil. Trans. Roy. Soc. (1855).
(8) de la Rive, ' Traite d'Electricite, ii. p. 776.
(9) Lodge, B. A. Report (1885) ; Phil. Mag. v. p. 19 (1885) ; xlix. pp. 351, 454 (1900).
(10) de Broglie, *Comp. Rend.* clii. p. 696 (1911).
(11) Sommerfeld, *Zeits. für Phys.* xlvii. p. 1 (1928).
(12) Dushman, Phys. Rev. xxi. p. 623 (1923).
(13) Becker and Brattain, Phys. Rev. xlv. p. 694 (1934).

XLIII. *Measurement of a Mutual Inductance and Resistance by a Ballistic Method.*

By E. TYLER, D.Sc., M.Sc., F.Inst.P., A.M.I.E.E., A.F.R.Ae.S.,
Head of the Science Department, South-East Essex
Technical College, Dagenham.*

[Received January 19, 1942.]

SUMMARY.

A ballistic galvanometer method, suitable for the measurement of a mutual inductance, is described.

The current energizing the primary winding of the inductance is produced by the repeated charge and discharge of a condenser. By effecting calibration of the galvanometer with the same capacitance, the mutual inductance is computed in terms of a known resistance and frequency of a vibrator-contactor. Various methods of operating and evaluating the frequency of the vibrator are also outlined, and comparison data for a pair of Pye coils reveal satisfactory consistency. Modifications of the method are indicated enabling an unknown resistance to be determined.

Introduction.

IN a recent article † by Astbury and Ford, a description is given of a ballistic method for measuring a capacitance C in terms of a mutual inductance M.

The method is one involving a comparison of consecutive ballistic deflections produced by (*a*) the passage of a quantity of electricity through a moving coil galvanometer by condenser discharge, (*b*) an induced charge flowing through the galvanometer as a result of reversing the steady current energizing the primary winding of the mutual inductance.

Although the method is accurate and possesses the advantage that no knowledge is required of the galvanometer constants, nor even of the resistance of the galvanometer circuit, computation of either M or C is not possible but only the ratio M/C. M may, however, be obtained independently of C if we resort to a multiple condenser charge-discharge method for supplying the steady primary current to the mutual inductance and calibrate the galvanometer for charge sensitivity with the same capacitance.

---

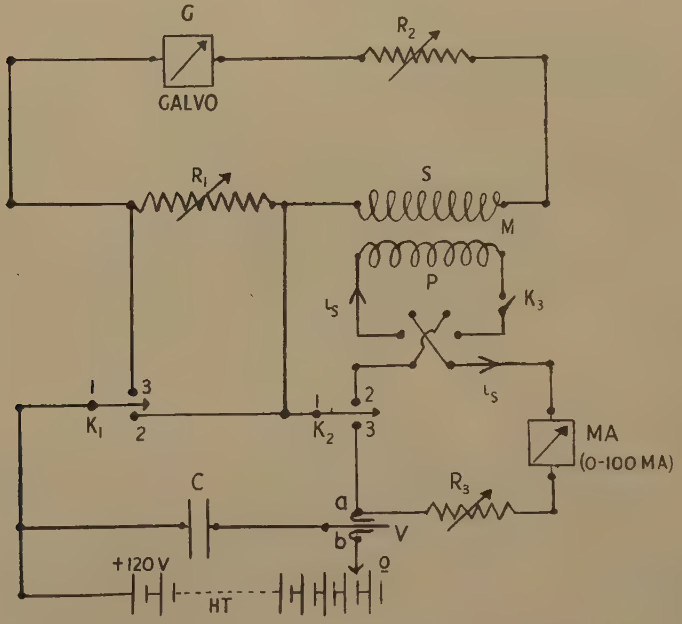
* Communicated by the Author.

† Journal Scientific Inst. vol. xv. no. 3 (March 1938).

Furthermore, in view of the theoretical result established by the authors in an appendix *, namely, the quantities of charge flowing through inductive resistances in parallel divide themselves inversely as the resistances, a ballistic galvanometer method is readily evolved for the measurement of an unknown resistance.

Descriptions of such ballistic methods suitable for the measurement of a mutual inductance and an unknown resistance are given herein.

Fig. 1.



C=Fixed condenser.
V=Electrically maintained make and break contactor.
G=Ballistic galvanometer.
M=Mutual inductance.

Method 1.—*Determination of a Mutual Inductance.*

The circuit employed for such an evaluation is shown in fig. 1.

A steady current i_s generated by the repeated discharge of a capacitance C, energizes the primary winding P of a mutual inductance M.

---

* *Loc. cit.* 1.

The establishment of such a current is effected by the operation of an electrically driven contact vibrator which periodically charges and discharges the condenser at a frequency N per sec., when connexions are made between the contacts 1 and 2 of the keys K_1 and K_2.

The potential V to which the capacitance is raised is derived from a super-capacity HT wireless battery (0–120 V.), so that the charge flowing per second through P=NCV, and is a measure of i_s, which, if so desired, may be registered by the milliammeter MA. In series with the ballistic galvanometer G is the secondary winding S of the mutual inductance, and two resistance boxes R_1 and R_2.

When the primary current i_s is established or reduced to zero by the key K_3, a ballistic deflexion D_1 is produced by virtue of the induced charge $\frac{M.i_s}{R_1+R_2+G+S}$ flowing through the galvanometer circuit, hence substituting NCV for i_s,

$$\frac{M.N.C.V}{R_1+R_2+G+S}=KD_1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (1)$$

K being the charge sensitivity constant of the galvanometer.

Elimination of K is effected by first disconnecting 1 and 2, then joining 1 and 3, and afterwards utilizing the vibrator as a single charge-discharge key. Thus, after charging C to a potential V, a charge from the condenser is sent through R_1 and R_2+G+S in parallel connexion.

The quantity of charge flowing through the galvanometer is therefore $\frac{C.V.R_1}{R_1+R_2+G+S}$, and this produces a ballistic deflection D_2 such that

$$\frac{C.V.R_1}{R_1+R_2+G+S}=KD_2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

In view of the resistance of the galvanometer circuit remaining fixed, K is the same as in the previous test, thus, elimination of K from equations (1) and (2) leads to the simple relation

$$M=\frac{R_1}{N}.\frac{D_1}{D_2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (3)$$

i.e., M is independent of C but dependent upon N, and N can be obtained accurately; hence the advantage of the method. For reversal of primary current

$$M=\frac{R_1}{2N}.\frac{D_1}{D_2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (4)$$

With the smaller M's in order to obtain suitable pairs of deflections of the same order it may be necessary to use different charging voltages

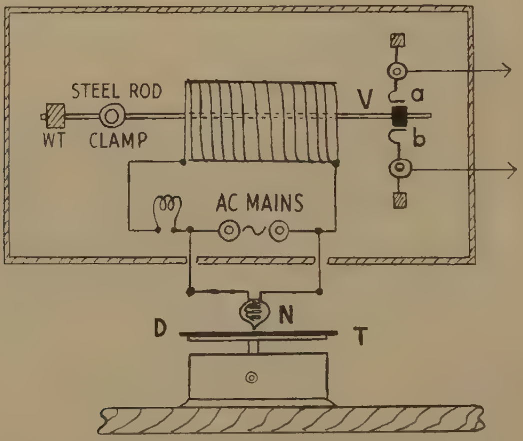
V_1 and V_2 to produce the deflections D_1 and D_2, in which case for make or break of primary current

$$M=\frac{R_1}{N}\cdot\frac{V_2}{V_1}\cdot\frac{D_1}{D_2}. \quad \ldots\ldots\ldots \quad (5)$$

The potentials are recorded by a reliable voltmeter, but when V_2 is small, a Rayleigh potentiometer suffices.

Furthermore, when testing with small M's, should the ballistic deflection due to the mutual inductance be small (2 or 3 cm. at normal working scale distance of 1 metre), increased accuracy is obtained by projecting the movement of the galvanometer light spot at normal working distance, upon a second distant screen at an increased linear magnification, observing on this screen the respective pairs of ballistic deflections. Alternately, by altering the series resistance R_2, corresponding pairs of D_1 and D_2 for any given value of R_2 are obtained, and plotting D_1 against D_2 produces a straight line graph the slope of which yields a more reliable value of $\frac{D_1}{D_2}$ for a given R_1.

Fig. 2 *a*.



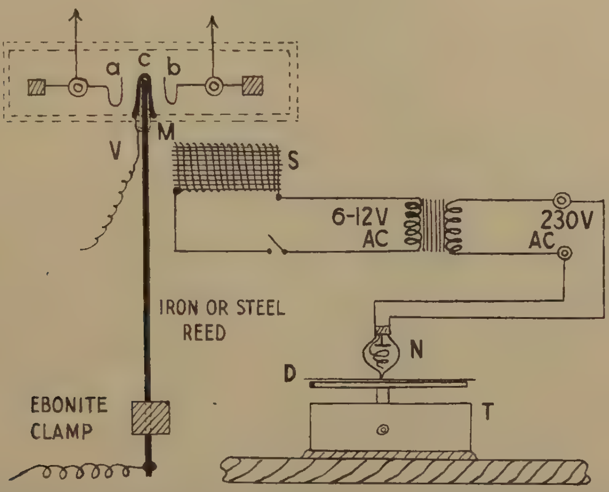
Microid vibrator contactor operated at 50 vibrations/sec.

V=Maintained vibrator contactor. T=Electric turntable.
N=Neon lamp. D=Stroboscopic disk.

Different types of vibrator contactors were used according to the frequency of operation. Whichever method is adopted, precision measurement of N is essential. In addition, it is imperative to insulate both the vibrator contactor and the fixed contacts from the remaining parts of the apparatus. Separate leads were taken from each contact and vibrator. At mains frequency of 50 cycles/sec. a steel rod incorporated in a "Microid Vibrator" * proved efficient (fig. 2*a*); likewise at 100 vibrations/sec. an electrically-driven soft iron reed tuned to resonance with the A.C. mains supply (fig. 2*b*). For frequencies below 50/sec. an electrically maintained steel reed † of variable but controlled length and with the usual mercury cup contact was used (fig. 3*a*).

A wooden lath clamped at one end and furnished with the necessary accessories also proved satisfactory (fig. 3*b*).

Fig. 2 *b*.



Electrically maintained reed contactor. Operated in resonance with the A.C. mains supply 50 cycles/sec.

V=Maintained vibrator.
M=Mica strips.
C=Copper contactor.
S=Soft iron core.

N=Neon lamp.
T=Electric turntable.
D=Stroboscopic disc.

For the higher frequencies above 100/sec. three arrangements were employed, viz. :—

(*a*) An electrically maintained fork loaded with similar contactors on each prong (fig. 4*a*).

(*b*) A power valve oscillator, driving a loud-speaker unit, the latter carrying the small steel vibrator (fig. 4*b*).

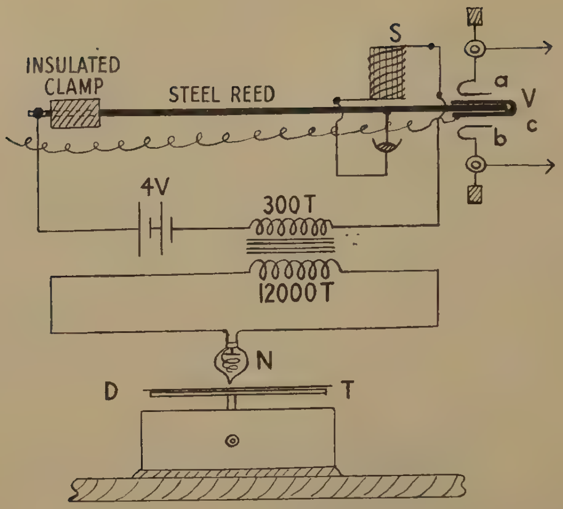
(*c*) A synchronous photo-electric cell amplifier device for energizing the same loud-speaker unit and reed as used in (*b*).

---

* Supplied by Griffin and Tatlock.

† Modified Hearson type.

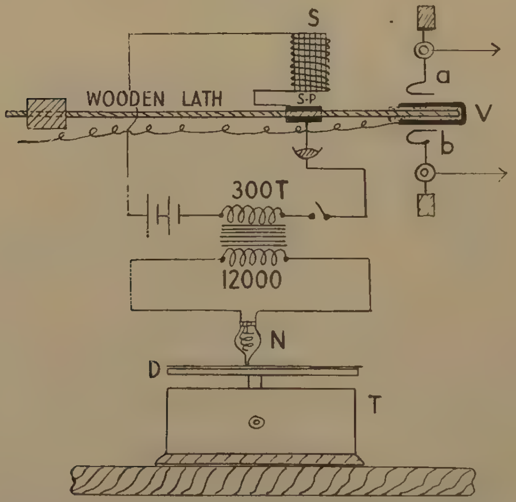
Fig. 3 *a*.


S
INSULATED
CLAMP
STEEL REED
a
V
c
b
4V
300T
12000T
N
D
T

Electrically maintained steel reed contactor (frequencies <50/sec.).

V=Maintained vibrator contactor carrying insulated copper contactor.
S=Soft iron electromagnet.
N=Neon lamp.
T=Turntable.
D=Stroboscopic disk.

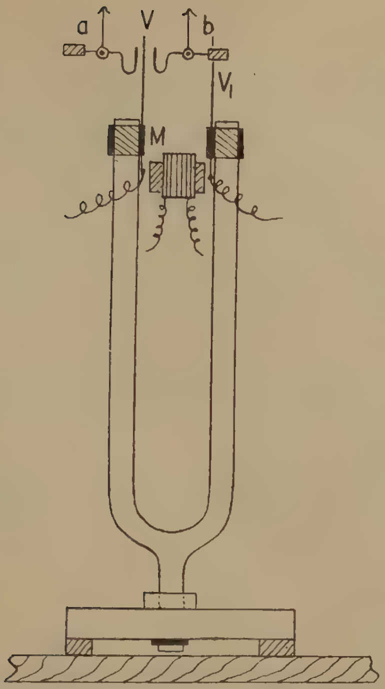
Fig. 3 *b*.


S
WOODEN LATH
S-P
a
V
b
300T
12000
N
D
T

Electrically maintained wooden lath carrying insulated contactor (frequencies <25/sec.).

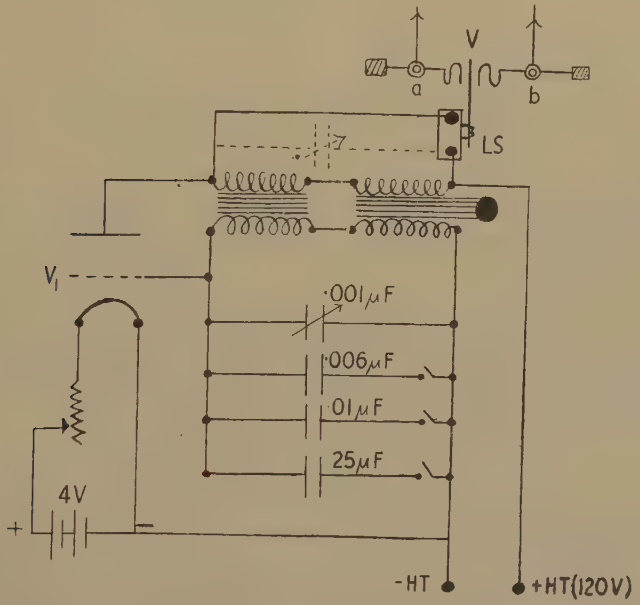
V=Maintained vibrator contactor insulated from lath.
SP=Soft iron collar on lath.
S=Electromagnet.
D=Stroboscopic disk.
N=Neon lamp.
T=Electric turntable.

Fig. 4 *a*.



Electrically maintained fork carrying insulated contactor V.
V_1=Dummy contactor. M=Mica, insulating V from fork.

Fig. 4 *b*.


V
a
b
LS
V1
·001μF
·006μF
·01μF
25μF
4V
+
−
-HT
+HT(120V)

Valve oscillator and loud-speaker unit operating vibrator contactor.
V_1=2 or 4 volt triode. LS=Loud-speaker unit driven by oscillator.
V=Insulated vibrator contactor, carried by loud-speaker unit.

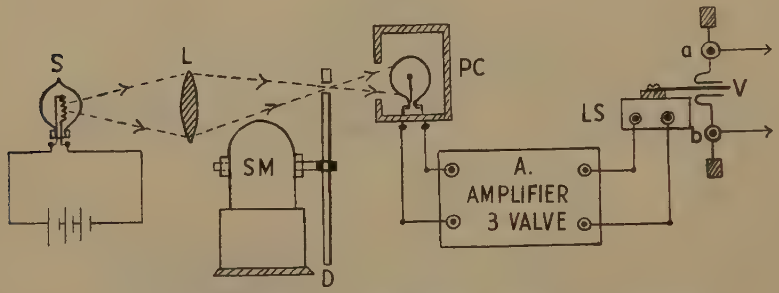
Since in methods (*b*) and (*c*) considerable energy is usually required to produce an appreciable amplitude of the vibrator, in order to enhance this production it is advantageous to operate the reed at resonance, by either adjusting the frequency of electrical circuit energizing the loud speaker or load the vibrator element with plasticine.

The frequency in each case was evaluated stroboscopically. A neon lamp (Osglim type) was made to glow intermittently by coupling it either directly or *viâ* a transformer to the energizing circuit operating the reed.

The intermittent light from the "Osglim" illuminated a stroboscopic disk with appropriate markings, the disk being rotated by an electric turntable when testing the frequency of the D.C. supply of that of the maintained reed or fork.

With the valve oscillator and photo-electric cell devices a synchronous motor operating from the A.C. mains supply was utilized. Thus, except in the cases of the electrical maintained reed and fork, we depend upon the frequency of the A.C. mains supply for our standard source. Tests over different long periods indicated this could be relied upon to 50±·02 cycles/sec.

Fig. 5.



A synchronous photo-cell amplifier device for operation of loud-speaker and vibrator contactor.

S=Light source (6 volt lamp).
L=Focusing lens.
PC=Photo-cell.
SM=Synchronous motor.

D=Slotted disk.
A=3-valve amplifier.
LS=Loud-speaker unit.
V=Insulated vibrator contactor.

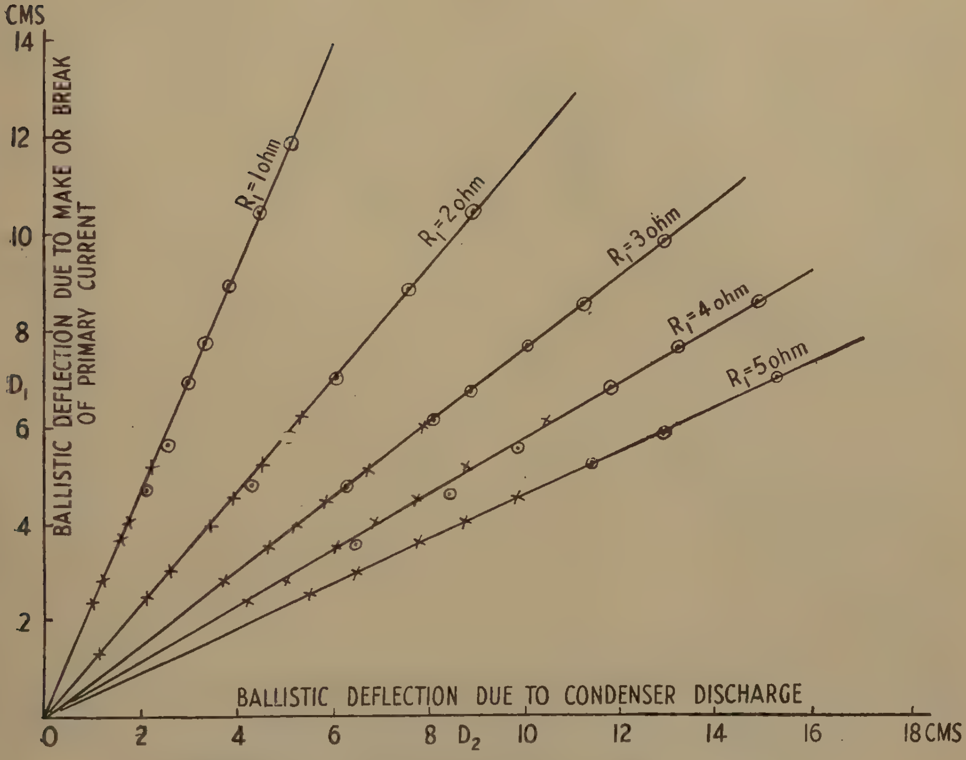
Specimen Results.

Referring to equations (3) and (5) we observe there are two ways of obtaining an accurate evaluation of M, viz. :—

(i.) by direct substitution of the appropriate data when D_1 and D_2 are of the same order ;

(ii.) by graphical representation when D_1 and D_2 are made variable. Since only under special conditions is $D_1 \simeq D_2$, we therefore resort to this method of analysis.

Fig. 6 *a*.



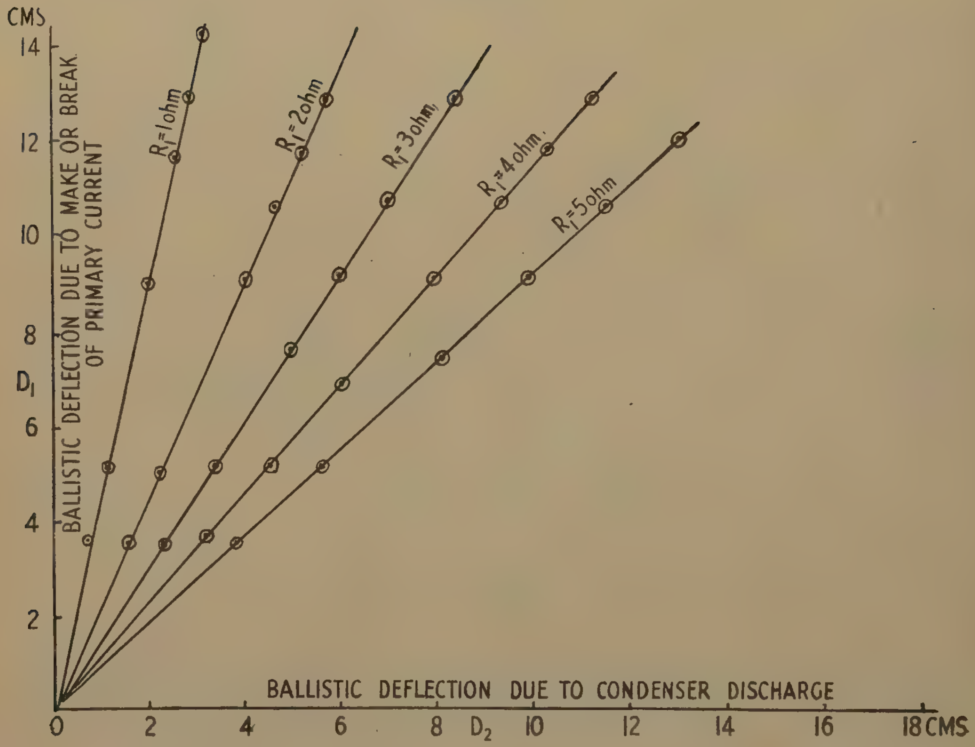
Results using the microid vibrator contactor and pair of Pye coils (500 turns each) closely coupled.

Charging voltage =110 volts.

⊙⊙⊙⊙⊙ $=C \simeq 4\mu F$. × × × × $=C \simeq 2\mu F$.

Frequency of vibrator=50/sec.

Fig. 6 *b*.



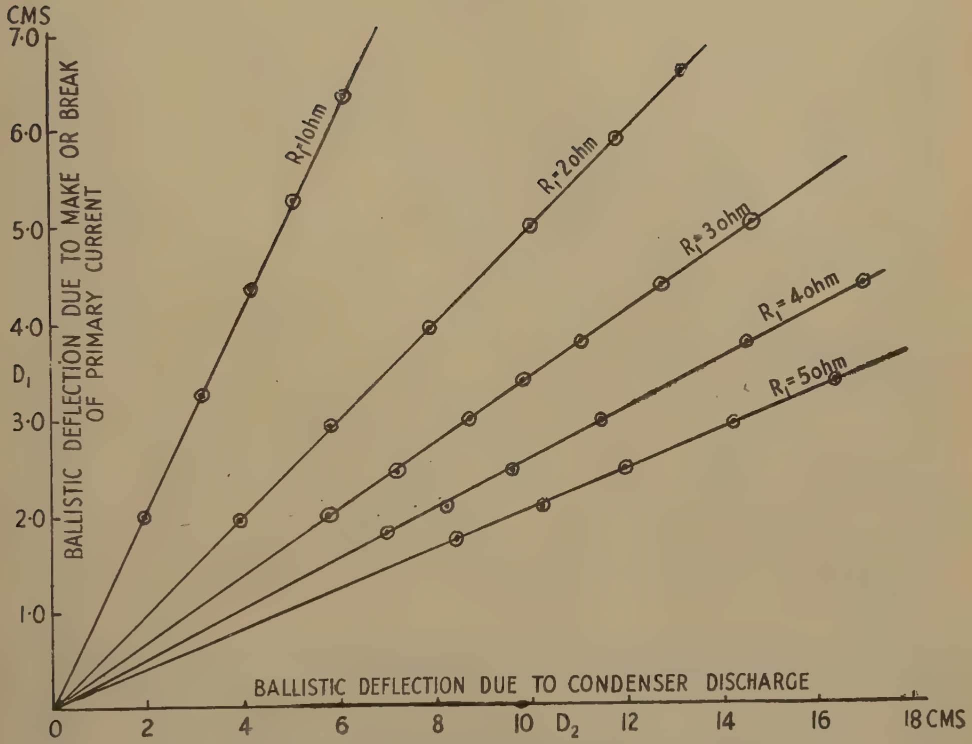
Results using " Resonant maintained iron reed contactor " and pair of Pye coils (500 turns each), closely coupled.

Charging voltage =109 volts.

⊙⊙⊙⊙⊙=$C \simeq 4\,\mu F$.

Frequency of vibrator=100/sec.

Fig. 7 *a*.



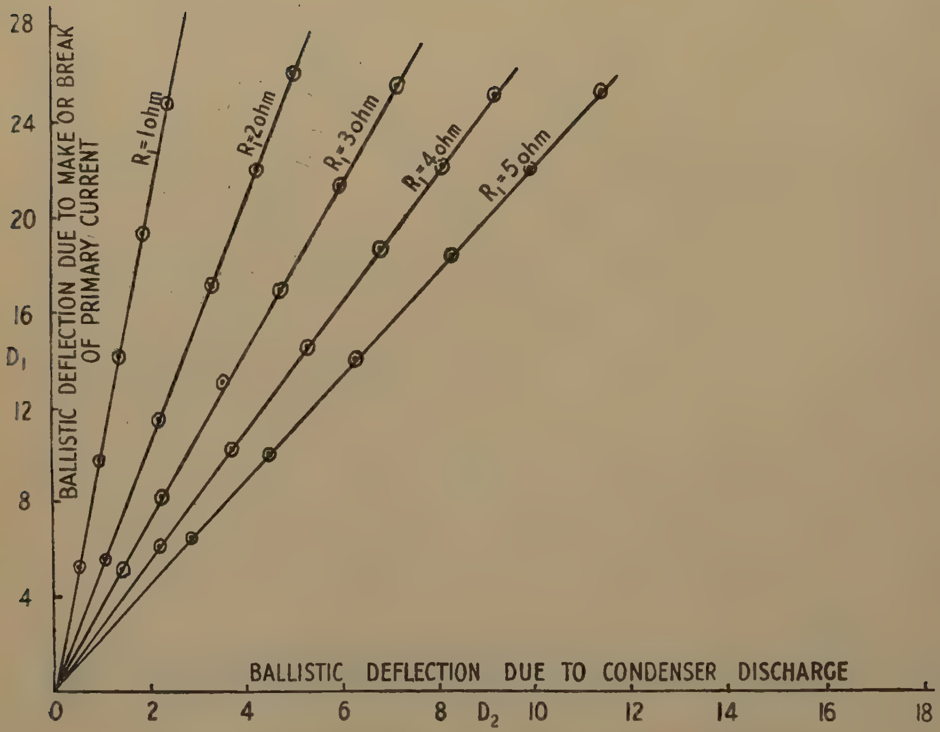
Results using the Hearson electrically maintained reed, and pair of Pye coils (500 turns each).

Charging voltage = 109 volts.

⊙⊙⊙⊙⊙ = $C \simeq 4\ \mu$F.

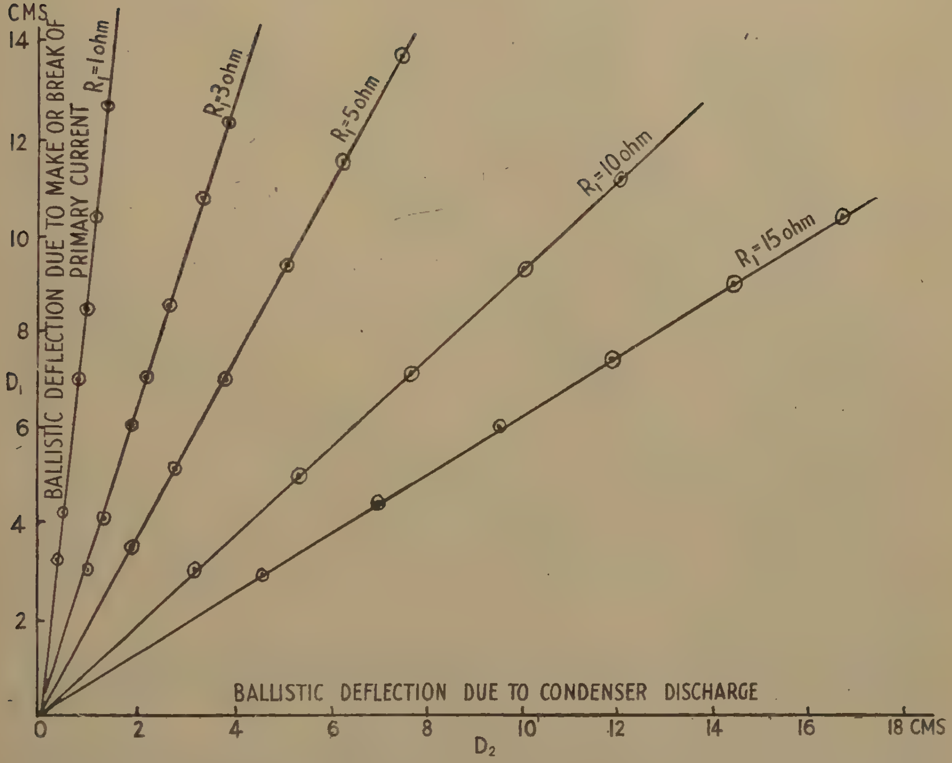
Frequency of vibrator = 22·31/sec.

Fig. 7 *b*.



Results using electrically maintained fork, and pair of Pye coils (500 turns each) closely coupled.
Charging voltage = 105 volts.
⊙⊙⊙⊙⊙ $= C \bumpeq 4\ \mu$F.
Frequency of fork = 241·3/sec.

Fig. 8 *a*.

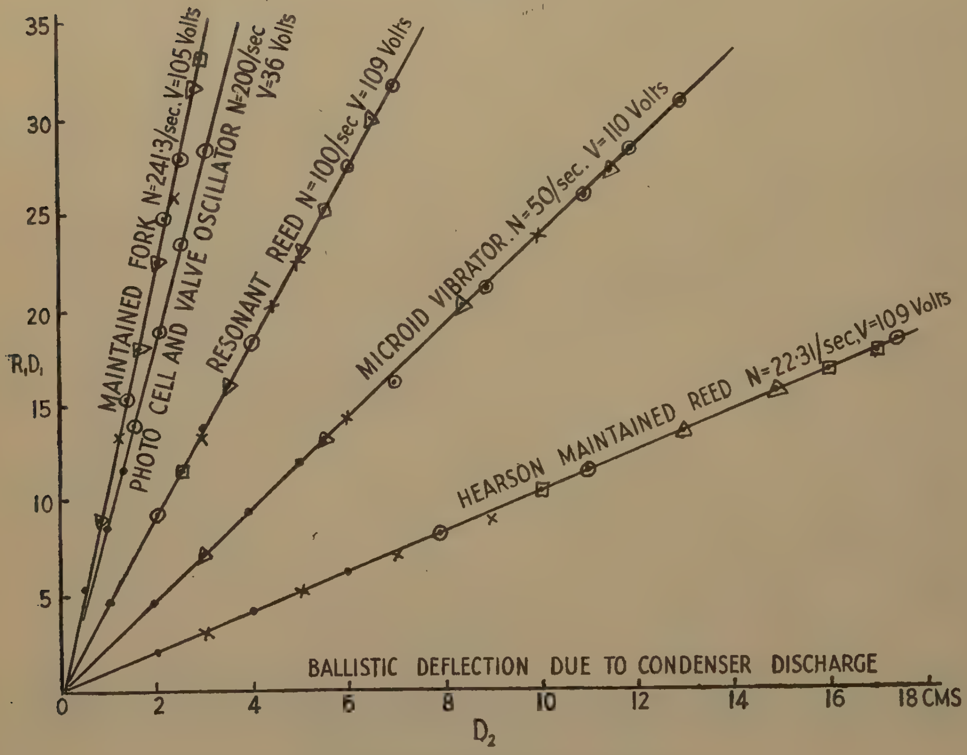


Results using valve oscillator or photo-cell-synchronous motor method with pair of Pye coils (500 turns each) closely coupled.

Charging voltage V=36 volts.
⊙⊙⊙⊙⊙=C≃4 μF.
Frequency of vibrator contactor=200/sec.

In order to obtain consecutive pairs of D_1 and D_2, R_1 is kept fixed and the resistance of the galvanometer circuit altered in stages by means of R_2. At each stage, comparison values of D_1 and D_2 are obtained. The tests are then repeated with other suitable fixed values of R_1.

Fig. 8 *b*.



Reduced graphs for each vibrator.

Typical results for each type of vibrator are shown in figs. 6*a*, 6*b*, 7*a*, 7*b* and 8*a*, where D_1 is plotted against D_2. The slope of each of the straight line graphs is compiled to give a mean value of $\frac{D_1}{D_2}$ for the known R's and N's, and substitution finally made for the derivation of M. (Table I.)

TABLE I.

Summary of Results from Figs. 6*a*, 6*b*, 7 *a*, 7*b*, and 8*a* for pair of Pye coils.
(500 turns each.) Diameter=15 cm. Maximum coupling.

Type of vibrator operating contactor.	Frequency of vibrator. N.	R_1	Slope of graph. $=\frac{D_1}{D_2}$	Mutual inductance. $M=\frac{R_1}{N}\cdot\frac{D_1}{D_2}$.
	vibs./sec.	ohms.		millihenries.
Microid, A.C. mains energized. (Fig. 2 *a*.)	50	1	2·34	46·8
		2	1·16	46·4
		3	·78	46·8
		4	·58	46·6
		5	·46 (5)	46·5
				Mean=46·5 (8)
Resonant soft iron reed. A.C. mains energized. (Fig. 2 *b*.)	100	1	4·65	46·5
		2	2·31	46·2
		3	1·54	46·2
		4	1·16 (5)	46·6
		5	·92 (5)	46·2 (5)
				Mean=46·3 (5)
Electrically maintained reed. (Hearson type.) (Fig. 3 *a*.)	22·3 (1)	1	1·04	46·6
		2	·51 (5)	46·1
		3	·34 (6)	46·6
		4	·26 (0)	46·6
		5	·20 (9)	46·7
				Mean=46·5 (2)
Electrically maintained fork. (Fig. 4 *a*.)	241·3	1	11·00	(45·5)
		2	5·62 (5)	46·5
		3	3·75	46·5
		4	2·78	46·0
		5	2·25	46·6
				Mean=46·4
Loud-speaker driven reed by photo-cell or valve oscillator method. (Figs. 4 *b* & 5.)	200	3	3·12 (5)	46·8 (7)
		5	1·85	46·2
		10	·92 (8)	46·4
		15	·62 (5)	46·8 (5)
				Mean=46·5 (6)

Plotting R_1D_1 against D_2 reduces the family of straight lines for a given vibrator at fixed frequency to one straight line of gradient $\frac{R_1D_1}{D_2}$ (fig. 8 *b*). This mean value of $\frac{R_1D_1}{D_2}$, together with a knowledge of N, is also employed to yield a more reliable value of M. (Table II.)

Table II.

Interpolated Data from Fig. 8 *b*.

Type of vibrator contactor.	Frequency of vibrator. N.	Slope of R_1D_1, D_2 graph.	Mutual inductance $M=\frac{\text{Slope of graph}}{N}$.
	vibs./sec.		millihenries.
Microid A.C. mains energized	50·0	2·33	46·5 (0)
Resonant soft iron reed. A.C. mains energized ..	100	4·64	46·4
Electrically maintained (Hearson type)	22·3 (1)	1·04	46·7 (0)
Electrically maintained fork	241·3	11·20	46·3 (0)
Loud speaker driven reed, by photo-cell or valve oscillator method	200	9·25	46·2 (5)
			Mean=46·4 (1)

Slope of graph fig. 9=M=46·4 (3) millihenries.
M, by Astbury & Ford's method=46·5 (4) ,,

A further plot of R_1D_1 against ND_2 for the various vibrators, reduces the family of straight lines of figs. 6 *a* to 8 *a* inclusive to one straight line of slope $\frac{R_1}{N}.\frac{D_1}{D_2}$ equal to M (fig. 9).

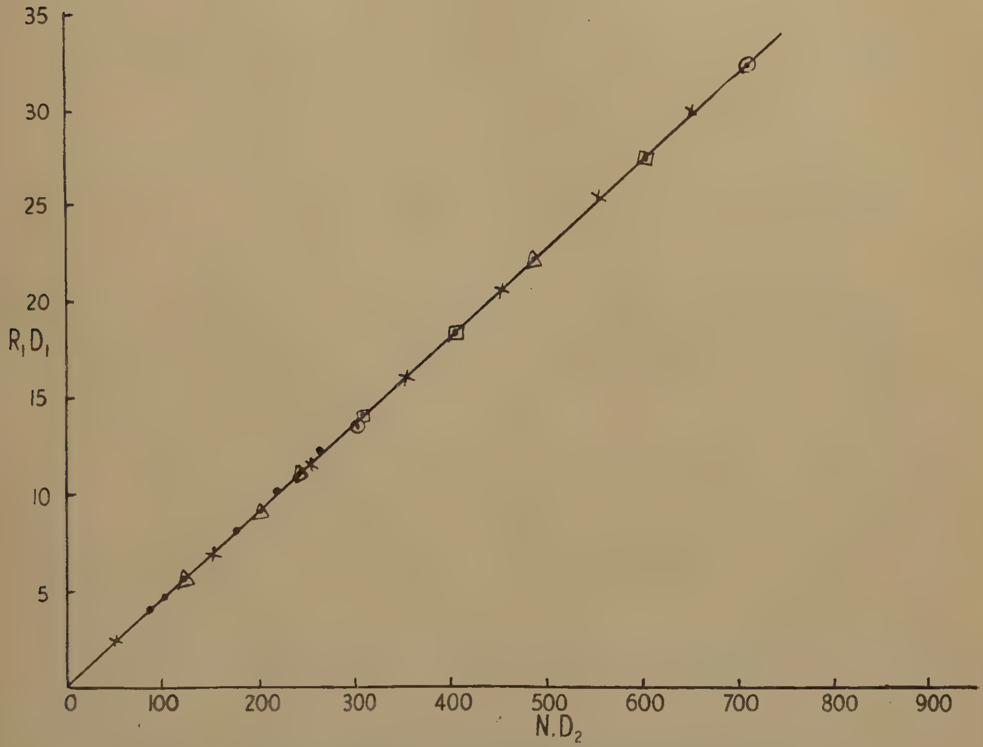
Deduced values of M for maximum coupling of a pair of Pye coils, each of 500 turns and mean diameter 15 cm., are included in Tables I. and II. The mean result 46·4 (3) millihenries compares well with an average value 46·5 (4) millihenries, obtained from several independent observations, employing Astbury and Ford's method.

Method 2.—Measurement of an Unknown Resistance.

The circuits shown in figs. 10*a* and 10*b* are employed, X being the unknown resistance and R a variable resistance box.

In both arrangements G+R+X is kept the same, while making comparison tests, thus ensuring constancy of the charge sensitivity of the galvanometer.

Fig. 9.



Reduced graph for different vibrators used in conjunction with two Pye coils (each 500 turns) closely coupled.

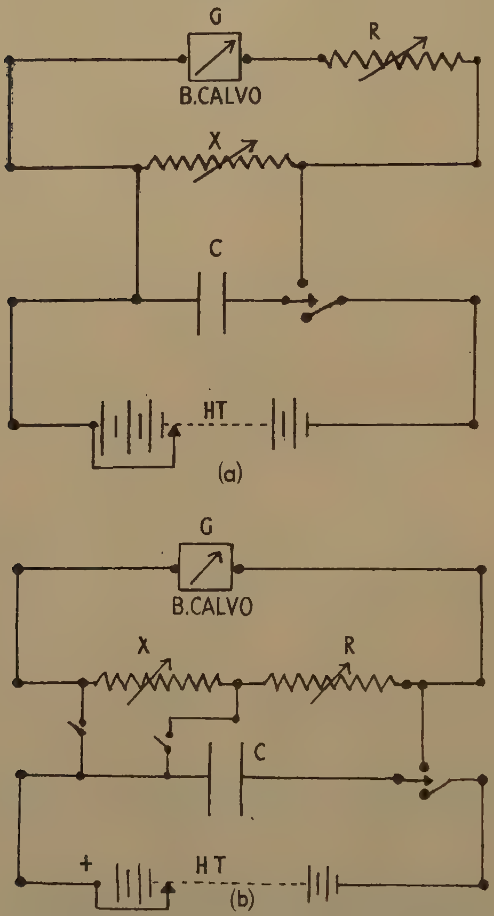
×	Microid vibrator.	N=50/sec.
⊙	Resonant reed.	N=100/sec.
..	Hearson reed.	N=22·31/sec.
△	Maintained fork.	N=241·3/sec.
⊡	Photo-cell and valve oscillator methods.	N=200/sec.

By interchanging X and R (fig. 10*a*) pairs of ballistic deflections, D_1 and D_2 are obtained for the discharges Q_1 and Q_2 through the galvanometer.

With the usual notation we thus have

$$Q_1=CV.\frac{X}{G+X+R}=KD_1, \quad \ldots \ldots \quad (6)$$

Fig. 10.



Ballistic method of comparing resistance.

and $$Q_2=CV\frac{R}{G+X+R}=KD_2, \quad \ldots \ldots \quad (7)$$

whereupon $$X=R.\frac{D_1}{D_2}. \quad \ldots \ldots \quad (8)$$

Thus by varying R we can obtain other values of $\frac{D_1}{D_2}$ and hence X.

It is preferable, however, to adjust both R and V so as to make D_1 the same order as D_2. A further arrangement outlined in fig. 10*b* may also be adopted.

Still keeping G+R+X fixed, R and R+X are successively placed across the condenser and the corresponding deflections D_2 and D_3 due to discharge observed. The charge Q_3 flowing through the galvanometer in this latter arrangement is expressible as

$$Q_3 = CV \,.\, \frac{R+X}{G+X+R} = KD_3\,; \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (9)$$

thus, combining expressions (7) and (9),

$$X = R \,.\, \frac{(D_3 - D_2)}{D_2}.$$

For this method to operate successfully $D_3 \simeq 2D_2$.

Table III.

Circuit used is shown in Fig. 10 *a*.
Capacity (C) = 2 μF.

Charging voltage V volts.	Resistance across condenser C.	Resistance in series with galvo G.	Ballistic deflections.		Unknown resistance $X=R.\frac{D_1}{D_2}$.
	ohm.	ohm.	D_1 cm.	D_2 cm.	ohm.
3	X= 40·0	R= 50·0	14·00 13·90 13·85 Mean=13·91		40·3 (1)
3	R= 50·0	X= 40·0		17·25 17·25 17·25 Mean=17·25	
3	X= 80·0	R=100	19·50 19·55 19·52 Mean=19·52		
3	R=100	X= 80·0		24·4 24·5 24·3 Mean=24·40	80·0

Specimen Results.

Typical data justifying the merits of these methods are included in Tables III. and IV. The unknown resistance X employed for checking purposes was an ordinary variable resistance box.

TABLE IV.

Circuit used is shown in fig. 10 *b*.
Capacity $=2\ \mu F$.

Charging voltage V.	Resistance across condenser C.	Resistance in series with galvo G.	Ballistic deflections.	Unknown resistance. $X=R.\frac{(D_3-D_2)}{D_2}$
Volts.	ohm.	ohm.		ohm.
3	R+X=60 (R=30, X=30)	0	D_3 23·6 23·5 23·5 23·6 Mean=23·55	30·12
3	R=30	X=30	D_2 11·8 11·7 11·7 11·8 Mean=11·75	
1·5	R+X=160 (R=80, X=80)	0	21·0 20·9 20·9 21·0 Mean=20·95	
1·5	R=80	X=80	D_2 10·45 10·50 10·45 Mean=10·46	80·2 (3)

XLIV. *Asymptotic Solutions of Linear Differential Equations.*

By Harold Jeffreys, F.R.S.*

[Received March 2, 1942.]

A type of approximate solution of certain types of differential equations, given by me in 1924, for application to Mathieu functions, has received considerable further application in quantum theory. The form that I gave was not the most convenient. The solutions can now be related directly to the Airy integral and its companion function, tables of which have been prepared by Dr. J. C. P. Miller and are in process of publication by the British Association. The present paper provides the direct connexion between the asymptotic solutions and these functions.

Many differential equations of physics are reducible to the form

$$\frac{d^2y}{dx^2}=\chi(x)y=(h^2\chi_0+h\chi_1+\chi_2)y, \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where h is large and independent of x, and χ_0, χ_1, χ_2 are given functions of x. Then in a range of x such that χ_0 does not vanish within it, asymptotic solutions for large h are

$$\chi_0^{-1/4}\exp\left[\pm\int^x(h\chi_0^{1/2}+\chi_1/2\chi_0^{1/2})dx\right]\{1+O(h^{-1})\}. \quad . \quad . \quad (2)$$

Similar results have been obtained by various authors (Horn 1899; Jeffreys, 1916; Fowler *et al.*, 1920; Wentzel, 1926; Brillouin, 1926), apparently originally by G. Green (1837), in showing that for tidal waves in a canal of slowly varying section the energy is transmitted without loss by reflexion (Lamb 1932). Carlini (1817) had given the Bessel Function case (*cf.* Watson 1922, p. 6). The present form is due to Jeffreys (1924). It is worth while to notice that the effect of χ_2 is of order h^{-1} and does not appear in the first term of the asymptotic expansion, while there are other terms of order h^{-1}. Consequently nothing is gained by including χ_2 in a first approximation, or by using such a form as $\chi^{1/4}\exp\{\pm\int\chi^{1/2}dx\}$ (Jeffreys 1936). In some cases this is known to be inferior in accuracy to (2)†. The type of case where the approxima-

---

* Communicated by the Author.

† For instance, the equation, originally suggested to me by Prof. G. H. Hardy,

$$y''-(h^2\theta'^2+h\theta'')y=0,$$

with $\theta=(\log x)^2$.

tion is useful is where χ varies greatly over a long range, but in such a way that its change in any range of order $\chi^{-1/2}$ is a small fraction of its mean value in that range ; in other words, if $\chi_0'/h\chi_0^{3/2}$ is small.

If χ_0 vanishes for some value of x within a range, this form of solution fails because the terms in h^{-1} contain negative powers of χ_0. But pairs of solutions of the form (2) will exist on both sides of the zero of χ_0. If this is a simple zero they will be of exponential type on one side and oscillatory on the other. It is a usual physical requirement that the solution on the exponential side shall be decreasing, and the problem is to determine the continuation of the same solution on the other side. This can be done as follows. We take χ_0 to vanish at $x=0$ and to be positive for $x>0$, and ignore χ_1. Then, on the negative side, we write $x=-\xi$, $\chi(x)=-\psi(\xi)$. Pairs of solutions are

$$\chi_0^{-1/4} \exp (\pm \mathrm{M}) \qquad x \geqslant 0, \qquad \qquad (3)$$

$$\psi_0^{\ 1/4} (\cos \mathrm{L}, \sin \mathrm{L}) \qquad x \leqslant 0, \qquad \qquad (4)$$

where

$$\mathrm{M}=\int_0^x h\chi_0^{1/2}dx \qquad \qquad (5)$$

$$\mathrm{L}=\int_0^\xi h\psi_0^{1/2}d\xi. \qquad \qquad (6)$$

If h is large and x small, the differential equation is approximately

$$\frac{d^2y}{dx^2}=h^2\chi_0'(0)xy, \qquad \qquad (7)$$

which is reducible to the equation satisfied by the Airy integral Ai(z).

In the British Association Tables in process of publication the function is defined in real form by

$$\mathrm{Ai}(x)=\frac{1}{\pi}\int_0^\infty \cos\left(\tfrac{1}{3}t^3+xt\right)dt,$$

and the companion solution by

$$\mathrm{Bi}(x)=\frac{1}{\pi}\int_0^\infty \left\{e^{-\frac{1}{3}t^3+xt}+\sin\left(\tfrac{1}{3}t^3+xt\right)\right\}dt.$$

The integral was introduced by Airy in an application to diffraction near a caustic. In the method of steepest descents, when the first approximation fails on account of the vanishing of the second derivative at the saddle-point, the next approximation involves the same integral. In the theory of dispersion it occurs in the neighbourhood of a stationary or limiting group-velocity (Jeffreys 1931).

If h is large enough there will be ranges on both sides of $x=0$, such that the conditions for the validity of (2) are satisfied, and, therefore, of (3) and (4), since the integral of the term in χ_1 will be small ; and also such that terms in $\chi_0''(0)x^2$ have not become large enough to produce serious error in (7). Thus the solutions in such ranges can be expressed in terms of the functions Ai and Bi, and comparison of these forms with (3) and (4) makes it possible to determine the combinations of asymptotic solutions on either side that correspond to the separate solutions on the other.

Put
$$z=\{h^2\chi_0'(0)\}^{1/3}x\,;\quad \zeta=-z. \qquad (8)$$

Then (7) becomes
$$\frac{d^2y}{dz^2}=zy, \qquad (9)$$

and the solutions are
$$y_1=\mathrm{Ai}(z)\,;\quad y_2=\mathrm{Bi}(z). \qquad (10)$$

Then for z large and positive
$$y_1\sim\frac{1}{2\sqrt{\pi}}z^{-1/4}\exp\left(-\tfrac{2}{3}z^{3/2}\right), \qquad (11)$$
$$y_2\sim\frac{1}{\sqrt{\pi}}z^{-1/4}\exp\left(\tfrac{2}{3}z^{3/2}\right), \qquad (12)$$

and for z large and negative
$$y_1\sim\frac{1}{\sqrt{\pi}}\zeta^{-1/4}\sin\left(\tfrac{2}{3}\zeta^{3/2}+\tfrac{1}{4}\pi\right), \qquad (13)$$
$$y_2\sim\frac{1}{\sqrt{\pi}}\zeta^{-1/4}\cos\left(\tfrac{2}{3}\zeta^{3/2}+\tfrac{1}{4}\pi\right). \qquad (14)$$

Also for $z>0$
$$\mathrm{M}\doteqdot\tfrac{2}{3}h\chi_0'(0)^{1/2}x^{3/2}=\tfrac{2}{3}z^{3/2}\,;\quad \chi_0^{-1/4}\doteqdot\{\chi_0'(0)x\}^{-1/4}=h^{1/6}z^{-1/4}\{\chi_0'(0)\}^{-1/6}, \qquad (15)$$
and for $z<0$
$$\mathrm{L}\doteqdot\tfrac{2}{3}\zeta^{3/2}\,;\quad \psi_0^{-1/4}=\{\chi_0'(0)\xi\}^{-1/4}=h^{1/6}\{\chi_0'(0)\}^{-1/6}\zeta^{-1/4}. \qquad (16)$$

Then the suitable pair of solutions for $z<0$ are
$$\left.\begin{aligned}\psi_0^{-1/4}\sin\left(\mathrm{L}+\tfrac{1}{4}\pi\right)&\sim h^{1/6}\pi^{1/2}\{\chi_0'(0)\}^{-1/6}y_1\\ &\sim\tfrac{1}{2}h^{1/6}\{\chi_0'(0)\}^{-1/6}z^{-1/4}\exp\left(-\tfrac{2}{3}z^{3/2}\right) \qquad z>0\\ &\sim\tfrac{1}{2}\chi_0^{-1/4}\exp(-\mathrm{M}),\end{aligned}\right\} \qquad (17)$$
and
$$\left.\begin{aligned}\psi_0^{-1/4}\cos\left(\mathrm{L}+\tfrac{1}{4}\pi\right)&\sim h^{1/6}\pi^{1/2}\{\chi_0'(0)\}^{-1/6}y_2\\ &\sim h^{1/6}\{\chi_0'(0)\}^{-1/6}z^{-1/4}\exp\left(\tfrac{2}{3}z^{3/2}\right) \qquad z>0\\ &\sim\chi_0^{-1/4}\exp(\mathrm{M}).\end{aligned}\right\} \qquad (18)$$

Hence asymptotic forms of the same solution, exponentially decreasing for $z>0$, are
$$\chi_0^{-1/4}\exp(-\mathrm{M})\longleftrightarrow 2\psi_0^{-1/4}\sin\left(\mathrm{L}+\tfrac{1}{4}\pi\right), \qquad (19)$$
and of the exponentially increasing function that resembles this in amplitude on the oscillatory side but differs in phase by $\tfrac{1}{2}\pi$,
$$2\chi_0^{-1/4}\exp\mathrm{M}\longleftrightarrow 2\psi_0^{-1/4}\cos\left(\mathrm{L}+\tfrac{1}{4}\pi\right), \qquad (20)$$
the pair on the left being valid for $x<0$ and those on the right for $x>0$.

The correspondence (19) was obtained by Jeffreys (1924) by expressing the solutions in terms of Bessel functions of order $\frac{1}{3}$. A result equivalent to (19) is given by Kramers (1926) (see also Kemble (1937), pp. 90–112). It is easier, however, to use the Airy integral directly, and Bessel functions of order $\frac{1}{3}$ seem to have no application except to provide an inconvenient way of expressing this function. The forms given by Jeffreys for the solution that increases for positive x do not correspond to the solution for negative x that has been chosen here as the most convenient. The difference in phase between (20) and equation (16) of Jeffreys's paper corresponds to the addition of a multiple of (19), which, however, is negligible in comparison with (20) when x is large and positive.

The condition for the validity of (3) and (4) is that $\chi_0'/h\chi_0^{3/2}$ shall be small; but in the range considered this is equal to $z^{-3/2}$, and the smallness of this is the condition that the asymptotic solutions shall be good approximations to the functions Ai(z) and Bi(z). Hence the conditions for the validity of the asymptotic solutions are equivalent, and the only further conditions that need attention are (1) that in the regions compared $\chi''(0)x^2$ shall be small compared with $\chi'(0)x$, and, therefore, that

$$\chi''(0) \ll h^{2/3}\chi'(0)^{4/3}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

and (2) that in the same regions the contribution from χ_1 shall be small, which gives

$$\chi_1(0) \ll h^{1/3}\{\chi'(0)\}^{2/3}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

The method was applied by N. F. Mott (1928) to derive the laws of classical statistical mechanics as approximations to those of quantum mechanics. He, however, quoted equation (16) of Jeffreys's paper instead of (15) or the above (19), having apparently overlooked the fact that the variable part of the M as there defined is negative, and the mistake was repeated in the first edition of Fowler's 'Statistical Mechanics' and apparently attributed to me in the second. The correct solution is given by Mott and Massey (1933). Kemble (1937) overlooks the fact that I gave also the correspondence for the second solution. The present L and M are defined to vanish at $x=0$, and otherwise to be positive in the respective ranges where they are used.

As an example of the method consider Bessel's equation for large order

$$x\frac{d}{dx}\left(x\frac{dy}{dx}\right)+(x^2-n^2)y=0, \quad . \quad . \quad . \quad . \quad . \quad (23)$$

and put

$$x=ne^{-z}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

Then

$$\frac{d^2y}{dz^2}=n^2(1-e^{-2z})y, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

the oscillatory side being for $x>n$ or $z<0$. Then for $x<n$

$$M=\int_0^z n(1-e^{-2z})^{1/2}dz=n(\theta-\tan\theta), \quad . \quad . \quad . \quad . \quad (26)$$

where $x=n \operatorname{sech}\theta$.

For $x>n$,

$$L=\int_0^\zeta n(e^{2\zeta}-1)^{1/2}d\zeta=n(\tan u-u), \quad . \quad . \quad . \quad . \quad (27)$$

where $x=n\sec u$. Then the correspondence from (19) is

$$\left.\begin{aligned} &2\tan^{-1/2}u\sin\{n\,(\tan u-u)+\tfrac{1}{4}\pi\}\\ &\quad\longleftrightarrow\tanh^{-1/2}\theta\exp\{-n(\theta-\tanh\theta)\}\\ &\quad=x^n\left(\frac{n^2-x^2}{n^2}\right)^{-1/4}\{n+(n^2-x^2)^{1/2}\}^{-n}\exp(n^2-x^2)^{1/2}, \end{aligned}\right\} \quad . \quad (28)$$

and the factor x^n identifies the solution as a constant multiple of $J_n(x)$. Comparison with the series solution for x small or the asymptotic solution for x/n large identifies the constant; our solution is $(2\pi n)^{1/2}J_n(x)$. For the other solution

$$\left.\begin{aligned} &2\tan^{-1/2}u\cos\{n\,(\tan u-u)+\tfrac{1}{4}\pi\}\\ &\quad\longleftrightarrow 2\tanh^{-1/2}\theta\exp n(\theta-\tanh\theta)\\ &\quad=2x^{-n}\left(\frac{n^2-x^2}{n^2}\right)^{-1/4}\{n+(n^2-x^2)^{1/2}\}^n\exp\{-(n^2-x^2)^{1/2}\}, \end{aligned}\right\} \quad (29)$$

which is the same constant multiple of $Y_n(x)$, apart from a change of sign. Apart from this factor the solutions are the Debye approximations obtained by the method of steepest descents (Watson 1922, pp. 241–5); but the present method has the advantage that it can be used in cases where convenient representations of the solutions by complex integrals are not known.

In some cases the argument is never large enough on the oscillatory side for the asymptotic approximations to be valid; but the Airy integral representation may still be valid on this side, and then the table of $\mathrm{Ai}(x)$ can be used directly. For Bessel functions of large order and argument this corresponds to the fact that if $x-n$ is of order $n^{1/3}$, $J_n(x)$ can be expressed in terms of Bessel functions of order $\frac{1}{3}$. Such a case has arisen in a problem of the surface waves of earthquakes, where variation of the properties with depth had to be taken into account (Jeffreys 1928). It was necessary there to have $\mathrm{Ai}'(x)/\mathrm{Ai}(x)$ to six places of decimals, since what was first determined was the wave-velocity, and the group-velocity had then to be found by numerical differentiation. In that paper x was taken with the opposite sign to that now recommended. It appears that in most problems where this type of approximation is useful the larger values of the argument, with the natural set of coordinates, are on the side where the solutions are exponential in character. For instance, in problems of atomic physics the natural argument is r, and the wave-functions are continuously decreasing for r more than a certain value and oscillatory for r less than that value. In problems of earthquake surface waves and some of electric waves in the atmosphere it is convenient to use the distance from the earth's surface as a coordinate, and then the solutions are oscillatory for small values of that coordinate and decreasing for large values. This does not hold for the usual Bessel functions, though it does hold if the order and argument are both imaginary, when the differential equation (23) is replaced by

$$x\frac{d}{dx}\left(x\frac{dy}{dx}\right)+(n^2-x^2)y=0.$$

In most cases some physical restriction excludes the solution Bi. But it can arise in some problems where the coefficient of y varies nearly linearly and the boundaries are separated by a finite amount, and we wish to make some allowance for the variation of the coefficient. Then, for instance, the differential equation may be put into the form

$$\frac{d^2y}{dx^2}=-k^3xy, \qquad \qquad (30)$$

with $y=0$ at $x=a$ and $x=b$. The solution is

$$y=\mathrm{A\ Ai}(-kx)+\mathrm{B\ Bi}(-kx), \qquad \qquad (31)$$

with

$$0=\mathrm{A\ Ai}(-ka)+\mathrm{B\ Bi}(-ka), \qquad \qquad (32)$$

$$0=\mathrm{A\ Ai}(-kb)+\mathrm{B\ Bi}(-kb), \qquad \qquad (33)$$

whence the characteristic equation determining k is

$$\mathrm{Ai}(-ka)\mathrm{Bi}(-kb)=\mathrm{Ai}(-kb)\mathrm{Bi}(-ka). \qquad \qquad (34)$$

References.

L. Brillouin, *Comptes Rendus*, 183, p. 24 (1926); *J. de Phys.* 7, p. 353 (1926).

R. H. Fowler, E. G. Gallop, C. N. H. Lock, and H. W. Richmond, Phil. Trans. A, 221, pp. 337–9 (1920).

G. Green, Camb. Phil. Trans. 6, pp. 457–62 (1837); Papers, p. 225.

A. Horn, *Math. Annalen*, 52, pp. 340–62 (1899).

H. Jeffreys, Memoirs R. A. S. 60, p. 211 (1916).

H. Jeffreys, Proc. Lond. Math. Soc. 23, pp. 428–36 (1924).

H. Jeffreys, M. N. R. A. S., Geophys. Suppl. 2, pp. 101–11 (1928).

H. Jeffreys, 'Operational Methods in Mathematical Physics' (1931).

H. Jeffreys, Phil. Mag. 21, pp. 544–6 (1936).

E. C. Kemble, 'Fundamental Principles of Quantum Mechanics' (1937).

H. A. Kramers, *Zs. f. Physik*, 39, p. 828 (1926).

H. Lamb, 'Hydrodynamics,' p. 274 (1932).

N. F. Mott, Proc. Camb. Phil. Soc. 24, pp. 76–9 (1928).

N. F. Mott and H. S. W. Massey, 'The Theory of Atomic Collisions,' p. 10 (1933).

G. N. Watson, 'Theory of Bessel Functions' (1922).

G. Wentzel, *Zs. f. Physik*, 38, p. 518 (1926).

XLV. *Differentially Cyclical Sets of Functions. An Extension of the Concept of Hyperbolic Functions.*

By LUDWIK SILBERSTEIN, Ph.D.*

[Received November 11, 1941.]

THE hyperbolic functions $f_1=\cosh x$, $f_2=\sinh x$ can be defined as a *cyclical dyad* of functions,

$$f_1'=f_2, \quad f_2'=f_1,$$

with $f_1(0)=1$, $f_2(0)=0$. The addition theorems of these functions and the algebraic relation between them, $f_1^2-f_2^2=1$, can all be derived from this definition without the aid of their power series expansions.

As a natural extension† of this dyad the concept of a *cyclical* n-ad suggests itself, *i. e.*, of a set of n functions f_k, such that

$$f_1'=f_2, \quad f_2'=f_3, \ldots f_n'=f_1.$$

Each function is then an integral of $d^n/dx^n=1$, and if we prescribe either initial values $f_1(0), \ldots f_n(0)$ or $f(0), f'(0), \ldots f^{(n-1)}(0)$ for one of them, the whole set is completely determined.

To fix the ideas, we shall claim

$$f_1(0)=1, \ f_1'(0)=f_1''(0)= \ldots =f_1^{(n-1)}(0)=0.$$

Then, necessarily,

$$f_2(0)=f_3(0)= \ldots =f_n(0)=0.$$

Also, $\sum_{k=1}^{n} f_k(x)=e^x$ (as in the case of the hyperbolic functions).

This follows at once from

$$\frac{d}{dx}\Sigma f_k=\Sigma f_k.$$

---

* Communicated by the Author.

† Having communicated the idea of such an extension and most of my results given in this paper to Prof. H. Bateman (Pasadena) towards the end of 1938, I have learned from him that Y. Villarceau, *C. R. Paris*, xc. (1880), deals with this subject and attributes the original idea of such an extension to the well-known Polish mathematician and mystic, Hoene Wroński. But I have, so far, no access to their writings.

An algebraic relation between the n functions of the set themselves is readily discovered. for any value of n. In fact, consider the determinant

$$D=\begin{vmatrix} f_1 & f_n & f_{n-1}\cdots f_2 \\ f_2 & f_1 & f_n \quad \cdots f_3 \\ \cdot & \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot & \cdot \quad\quad \cdot \\ f_n & f_{n-1} & f_{n-2}\cdots f_1 \end{vmatrix}.$$

The derivative dD/dx is the sum of the n determinants obtained from D by placing dashes on the elements of the first, or the second, etc., the nth row. But thus the first row becomes

$$f_1', \quad f_n', \ldots f_2'=f_2, \quad f_1, \ldots f_3,$$

identical with the second row, and similarly for each of the remaining determinants. Hence, $dD/dx=0$, $D=\text{const.}$, and since $D(0)=1$, we have

$$\begin{vmatrix} f_1 & f_n & \cdots f_2 \\ f_2 & f_1 & \cdots f_3 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ f_n & f_{n-1} & \cdots f_1 \end{vmatrix}=1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (A)$$

which is the required relation between the n functions of the set. *E.g.*, for $n=2$, $\begin{vmatrix} f_1 & f_2 \\ f_2 & f_1 \end{vmatrix}=1$, the familiar relation between the hyperbolic functions, $\cosh^2 x-\sinh^2 x=1$.

The next question is : Are there, as in the case of $n=2$ and of $n=1$, *addition theorems* for the functions of a cyclical set, and what is their form ? And, finally : What are the relations between $f_k(-x)$ and $f_1(x), \ldots f_n(x)$?

The method of finding the addition theorems and the last-mentioned relations will be sufficiently explained by considering the case of a *cyclical triad* of functions.

Thus, let

$$f'(x)=g(x), \quad g'(x)=h(x), \quad h'(x)=f(x) \quad . \quad . \quad . \quad . \qquad (1)$$

and

$$f_0=1, \quad f_0'=0, \quad f_0''=0.$$

Then, necessarily,

$$g_0=0, \quad g_0'=0, \quad g_0''=1,$$

$$h_0=0, \quad h_0'=1, \quad h_0''=0,$$

and the three functions, each a solution of $\frac{d^3}{dx^3}=1$, are herewith completely determined *. But in deriving the announced properties of the functions f, g, h, we shall not use their expansions.

Let x and y be two independent arguments. What are $f(x+y)$, etc., in terms of the three functions of x and of y themselves? Let $F(x, y)$ be any aggregate of the latter functions. The necessary and sufficient condition that F shall depend on

$$u=x+y$$

alone is $\partial F/\partial x=\partial F/\partial y$. Thus, to construct the required expression, start with

$$F=f(x)\,f(y)+\ldots,$$

to be properly supplemented. This gives

$$\frac{\partial F}{\partial x}=g(x)f(y)+\ldots, \quad \frac{\partial F}{\partial y}=f(x)\,g(y)+\ldots.$$

Thus we must add $f(x)g(y)$ to the former and $g(x)f(y)$ to the latter derivative, and therefore, to F itself the two terms $h(x)g(y)$ and $g(x)h(y)$. This gives

$$F=f(x)f(y)+g(x)h(y)+h(x)g(y),$$

which turns out already to have the required property $\partial F/\partial x=\partial F/\partial y$. Thus, $F=F(u)$.

Moreover, since $dF/du=\partial F/\partial x$, we have

$$F'(u)=g(x)f(y)+h(x)h(y)+f(x)g(y),$$

$$F''(u)=h(x)f(y)+f(x)h(y)+g(x)g(y),$$

and $F'''(u)=F(u)$; and since also $F_0=1$, $F_0'=0$, $F_0''=0$, it follows that $F(u)=f(u)$.

This, then, gives an expression for $f(x+y)$, whence the remaining two expressions follow by differentiation. In fine,

$$\left.\begin{aligned} f(x+y)&=f(x)f(y)+g(x)h(y)+h(x)g(y),\\ g(x+y)&=h(x)h(y)+f(x)g(y)+g(x)f(y),\\ h(x+y)&=g(x)g(y)+h(x)f(y)+f(x)h(y). \end{aligned}\right\} \quad \ldots\ldots \quad (2)$$

---

* In fact,

$$f=1+\frac{x^3}{3\,!}+\frac{x^6}{6\,!}+\ldots=\frac{1}{3}\left[e^x+2e^{-\frac{x}{2}}\cos\frac{\sqrt{3}}{2}x\right],$$

$$g=\frac{x^2}{2\,!}+\frac{x^5}{5\,!}+\ldots=f'=\frac{1}{3}\left[e^x-2e^{-\frac{x}{2}}\cos\left(\frac{\sqrt{3}}{2}x-\frac{\pi}{3}\right)\right],$$

$$h=x+\frac{x^4}{4\,!}+\ldots=g'=\frac{1}{3}\left[e^x-2e^{-\frac{x}{2}}\cos\left(\frac{\sqrt{3}}{2}x+\frac{\pi}{3}\right)\right].$$

For the first of these expressions see Phil. Mag. xxv. p. 1004 (1938).

These are the three *addition theorems* of our cyclical triad of functions. They are, comparatively, not more complicated than those for the hyperbolic functions. Instead of sums of two binary products we have here sums of three such products.

In particular, for $x=y$,

$$\left.\begin{aligned} f(2x)&=f^2(x)+2g(x)h(x), \\ g(2x)&=h^2(x)+2f(x)g(x), \\ h(2x)&=g^2(x)+2h(x)f(x). \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (2.1)$$

The algebraic relation between the three functions is, by (A),

$$\begin{vmatrix} f & h & g \\ g & f & h \\ h & g & f \end{vmatrix} = 1, \quad \ldots\ldots\ldots \quad (3)$$

or, in developed form,

$$f^3+g^3+h^3-3fgh=1. \quad \ldots\ldots\ldots \quad (3')$$

This is, unlike the hyperbolic relation, perfectly symmetrical in f, g, h*. There is, of course, one more relation between the three functions, but it is not algebraic. Using, *e. g.*, the first of the formulæ given in footnote to page 459, we may put the second relation into the form

$$2f-g-h=\frac{2}{\sqrt{f+g+h}}\cos\frac{\sqrt{3}}{2}\log(f+g+h).$$

The elimination of $\cos\frac{\sqrt{3}}{2}x$, $\sin\frac{\sqrt{3}}{2}x$ between any two of the expressions for f, g, h gives only again the relation (3′).

To express $f(-x)$, etc., in terms of $f(x)$, etc., apply (2) to the case $x+y=0$. This gives

$$\begin{aligned} f(x)f(-x)+h(x)g(-x)+g(x)h(-x)&=1, \\ g(x)f(-x)+f(x)g(-x)+h(x)h(-x)&=0, \\ h(x)f(-x)+g(x)g(-x)+f(x)h(-x)&=0, \end{aligned}$$

three linear equations for $f(-x)$, etc. The determinant of this system is, by (3), equal to 1. Thus

$$\left.\begin{aligned} f(-x)&=f^2(x)-g(x)h(x), \\ g(-x)&=h^2(x)-f(x)g(x), \\ h(-x)&=g^2(x)-h(x)f(x), \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (4)$$

---

* The first member of the equation (3′) has the factor $f(x)+g(x)+h(x)$, the other factor being $f^2+g^2+h^2-gh-hf-fg$, or, by (4) *infra*, $f(-x)+g(-x)+h(-x)$. As we saw, the first factor is e^x, and the second, e^{-x}, which only verifies the relation (3′) ; $e^xe^{-x}=1$.

which are the required relations. (They follow also from each other by differentiation.) These relations are naturally more complicated than those for the hyperbolic functions, since none of our three functions is even or odd.

This completes, essentially, the theory of the cyclical triad of functions. The triad has seemed particularly interesting in view of the comparative simplicity of its addition theorems and the symmetry of the algebraic relation (3′). It may, perhaps, find applications in connexion with problems in a three-space. The subject has, at any rate, seemed interesting for its own sake.

A table and graphs of the functions f, g, h may be easily constructed by means of their expressions given in the footnote to page 459. The argument x of these functions need not, of course, be limited to real values, but may be extended to the complex domain.

The details concerning the cyclical tetrad and higher sets of functions will be left to the reader. I may mention only that for a cyclical tetrad. f, g, h, k, the addition theorems are

$$f(x+y)=f(x)f(y)+h(x)h(y)+g(x)k(y)+k(x)g(y),$$

and three more following therefrom by differentiation. Two relations between the four functions are

$$(f+h)^2-(g+k)^2=1, \quad (f-h)^2+(g-k)^2=1.$$

But one of these follows from the other in view of (3).

759 Flower City Park,
Rochester, N.Y.
September 28, 1941.

XLVI. *Eigenfunctions in Heisenberg approximation of the Two-electrons Problem.*

By L. Pincherle, King's College, London,
at the H. H. Wills' Laboratory, University of Bristol *.

[Received January 19, 1942.]

1. *Introduction.*

As is well known, Heisenberg's approximation for the helium problem introduces a potential V which is symmetric in the co-ordinates of the two electrons (and so enables all approximations of Schrödinger perturbation theory to be carried out), namely † :—

$$V=V(r_1)+V(r_2)\ ;$$

$$V(r)=-\frac{Ze^2}{r}+\frac{e^2}{r_0},\quad \text{when } r<r_0,$$

$$V(r)=-\frac{(Z-1)e^2}{r},\quad \text{when } r>r_0$$

(e is the charge of the electron and Z the nuclear charge in e units). r_0 must be so chosen that on the average one of the electrons is inside and the other outside the sphere of radius r_0.

This approximation has been used especially in cases in which one of the electrons is almost constantly nearer to the nucleus than the other one, so that at the former there is almost exactly the potential $\frac{Ze^2}{r}+\frac{e^2}{r_0}$ and at the latter the potential $-\frac{(Z-1)e^2}{r}$.

It is the purpose of this note to give the analytical expression of the eigenfunctions of the single electrons in Heisenberg's approximation, and to apply it to a case in which the distance of the two electrons from the nucleus is the same.

2. *Analytical Expression of the Eigenfunctions.*

The differential equation (of the confluent hyper-geometric type ‡) satisfied by the radial eigenfunction ψ of a single electron (we denote by ψ the ordinary eigenfunction multiplied by r, so that $\int_0^\infty \psi^2 dr=1$) is, in atomic units,

---

* Communicated by Dr. H. T. Flint.

† W. Heisenberg, *Zeit. Phys.* xxxix. p. 499 (1927).

‡ See Whittaker and Watson, 'Modern Analysis,' Cambridge, 1927, Chapter XVI.

$$\frac{d^2\psi_i}{dr^2}+\left[-\alpha^2-\frac{2}{r_0}+\frac{2Z}{r}-\frac{l(l+1)}{r^2}\right]\psi_i=0$$

in the inside of the sphere of radius r_0, and

$$\frac{d^2\psi_e}{dr^2}+\left[-\alpha^2+2\frac{(Z-1)}{r}-\frac{l(l+1)}{r^2}\right]\psi_e=0$$

outside the same sphere.

l is the azimuthal quantum number, and $-\alpha^2=2E$, where E is the energy. The boundary conditions are the usual ones, namely ψ must be finite and continuous everywhere, and must vanish for $r=0$ and $r=\infty$; for $r=r_0$ we must have

$$\psi_i(r_0)=\psi_e(r_0)\,;\quad \frac{d\psi_i}{dr}(r_0)=\frac{d\psi_e}{dr}(r_0)\,;$$

moreover, as the solution must represent an electron in a given quantum state n, ll, it must have, between $r=0$ and $r=\infty$, the right number of zero points, namely $n-l-1$.

The solution for ψ_i can conveniently be written in the following form :—

$$\psi_i=Cr^{l+1}e^{-Kr}F\left[l+1-\frac{Z}{K},\ 2l+2,\ 2Kx\right],$$

where

$$F(a, b, x)=1+\frac{a}{b}x+\frac{a(a+1)}{b(b+1)}\frac{x^2}{2!}+\ldots,\quad \text{and}\quad K=\sqrt{\alpha^2+\frac{2}{r_0}}.$$

For $r_0\to\infty$ it becomes the ordinary solution of the one-electron problem for a nuclear charge $=Z$.

The solution for ψ_e can be expressed as follows :—

$$\psi_e=C'e^{-\alpha r}r^\beta G(r),\quad \text{with}\quad \beta=\frac{Z-1}{\alpha},$$

where we can put
$$G(r)=\sum_{n=0}^{\infty}c_n r^{-n}.$$

The c's satisfy the following recurrent relation :—

$$c_0=1,$$
$$-2\alpha c_1=\beta(\beta-1)-l(l+1),$$
$$-2\alpha n c_n=[(\beta-n+1)(\beta-n)-l(l+1)]c_{n-1},$$

which shows that G(r) is a half-convergent series (analogous, for instance, to the series which expresses the error function for large values of the argument). The present form of the solution can then be used only when r_0 is big enough for the first few terms of the half-convergent series to give a sufficient approximation to the value of $G(r_0)$. This, however, is the case in most applications.

3. *Determination of E and* r_0.

In the formulæ of the preceding section we have, for each electron, four parameters to be determined, namely, the two constants C and C′, the eigenvalue E (or α), and r_0, which is, however, the same for the two electrons. A practical way of determining them is the following :—we consider one of the electrons and fix arbitrarily a value of the energy, say E′ (that must, of course, be intermediate between the known values that E should have for r_0 $r=0$ and $\overline{r_0}$ $r=\infty$), and a value of r_0, say $\overline{r_0}$. We determine C′, as a function of C, from the condition $\psi_i(\overline{r_0})=\psi_e(\overline{r_0})$. Then, generally, $\frac{d\psi_i}{dr}(\overline{r_0})$ is not equal to $\frac{d\psi_e}{dr}(\overline{r_0})$, but we can vary $\overline{r_0}$ until this relation is satisfied. We have then a couple of values, E′, $\overline{r_0}$, for which the solution $\psi_2(E', \overline{r_0})$, satisfying all the conditions, is known ; the constant C is determined by the normalization condition $\int_0^\infty \psi^2\,dr=1$. (This integration, and others which follow, must be carried out numerically.) It is more convenient to fix E′ and vary $\overline{r_0}$ than to do the opposite, because $G(r)$ is independent of r_0 and then the coefficients c_n have not to be calculated again at each attempt.

Now the procedure must be repeated for the second electron ; only this time $\overline{r_0}$ is fixed, as it must be the same for the two electrons, and we must then vary E until we find a value E″ for which the solution with the right number of zero points is reached and the condition $\frac{d\psi_i}{dr}(\overline{r_0})=\frac{d\psi_e}{dr}(\overline{r_0})$ is satisfied. Let us call this solution $\psi_2(E'', \overline{r_0})$. We have now eigenvalues and eigenfunctions of the two electrons corresponding to the particular value $\overline{r_0}$. If, now, on the average one of the electrons is found inside and the other outside the sphere of radius $\overline{r_0}$, namely, if

$$\int_0^{r_0} \psi_1^2\,dr+\int_0^{r_0} \psi_2^2\,dr=1,$$

we have completely solved our problem. But this is not generally the case at the first attempt, and other particular values of r_0 must be tried until the last condition is satisfied.

The procedure may appear cumbersome, but is not really very long, because the physical conditions of the problem enable us to limit the values of all the parameters within a narrow range.

4. *Application to the $(2p)^2$-state of Helium.*

The procedure explained in section 3 becomes simpler when the two electrons are in the same state. In fact, in this case both electrons are represented by the same eigenfunction $\psi(E', \overline{r_0})$, and so it is only necessary to find the couple of values E', $\overline{r_0}$ which satisfy the conditions explained in section 3, and for which $\int_0^{\overline{r_0}} \psi^2 dr = \frac{1}{2}$.

This method has been applied to the (virtual) $(2p)^2$-state of helium.

The values of E and r_0, for which $\int_0^{r_0} \psi^2 dr = \frac{1}{2}$, are found to be :

$E' = -0{\cdot}3673$ Rydberg ; $\overline{r_0} = 2{\cdot}843$ (the radius of the first Bohr hydrogen orbit being the unit of length). The analytic expression of the eigenfunction is :

$$\psi_i = 0{\cdot}9476 e^{-1{\cdot}0348 r} r^2 F[0{\cdot}067256\ ;\ 4\ ;\ 2{\cdot}0696 r] \quad (r < 2{\cdot}843).$$

$$\left.\begin{aligned} \psi_e &= 0{\cdot}3729 e^{-0{\cdot}60605 r} r^{1{\cdot}6500} G(r) \\ G(r) &= 1 + \frac{0{\cdot}7652}{r} + \frac{0{\cdot}7031}{r^2} + \frac{0{\cdot}2953}{r^3} - \frac{0{\cdot}0714}{r^4} + \dots \end{aligned}\right\} (r > 2{\cdot}843).$$

TABLE I.

Value of	Calculated with			Experi-mental.
	Present eigenf.	Hartree eigenf.	Hydrogen eigenf.	
E'	$-0{\cdot}3673$	$-0{\cdot}386$	—	—
F_0	$0{\cdot}586$	$0{\cdot}576$	—	—
W	$-1{\cdot}321$	$-1{\cdot}348$	$(-1{\cdot}348)$	$-1{\cdot}368$
$\bar{r}$	$3{\cdot}15$	$3{\cdot}18$	$3{\cdot}04$	—
r^2	$12{\cdot}54$	$12{\cdot}57$	$11{\cdot}13$	—

The value of the $(2p)^2$-term is obtained, as in the Hartree approximation, subtracting from the sum of the eigenvalues of the two electrons, $2E'$, their electrostatic interaction

$$F_0 = \iint_0^\infty \psi^2(r_1) \frac{e^2}{r_{12}} \psi^2(r_2)\, dr_1\, dr_2$$

(where r_{12} is the distance of the two electrons).

The value of F_0 results, $F_0 = 0{\cdot}586$ Rydberg, so that the term value is $W = -2.0{\cdot}3673 - 0{\cdot}586 = -1{\cdot}321$ Rydberg.

Table I. gives the values of E' and F_0, the value of W (the $(2p)^2$-term), and the value of $\bar{r}$ (average distance of the electrons from the nucleus) and $\bar{r^2}$ (average of the square of the same quantity), calculated with the present eigenfunctions, and with Hartree eigenfunctions *. The two last quantities have also been calculated with a hydrogen eigenfunction in which an effective nuclear charge has been introduced which gives the same term-value as the self-consistent field method. An experimental value of the $(2p)^2$-term is also reported, which is deduced from an observation by Kruger † of the $(2p)^2$ ^3P-term, but this value is uncertain. A calculation by Ta-You Wu ‡, using a modified form of the variational method, gives for the $(2p)^2$-term the value $-1{\cdot}331$.

The method has also been applied to the fundamental state of helium, though in this case the first few terms of the semi-convergent series do not give a very good approximation to the value of $G(r_0)$. The value 23·29 volt has been found for the ionization potential of He, a value near to that, 23·46, obtained with the Hartree method ; the experimental value is 24·47.

We see that the present eigenfunction, though less accurate than Hartree's, give results in the two-electrons problem not very different from it, while, on the other hand, having the advantage of allowing all approximations of the perturbation theory to be carried out, and of being represented by simple functions.

Summary.

The analytical expression of the eigenfunctions in Heisenberg's approximation of the two-electrons problem is given, and an application made to the $(1s)^2$— and $(2p)^2$— states of helium.

---

* W. S. Wilson and R. B. Lindsay, Phys. Rev. xlvii. p. 681 (1935).

† P. G. Kruger, Phys. Rev. xxxvi. p. 855 (1930).

‡ Ta-You Wu, Phys. Rev. xlvi. p. 239 (1934).

XLVII. *The Drifting Force on a Ship among Waves.*

By T. H. Havelock, F.R.S.*

[Received March 9, 1942.]

1. When a ship is advancing through a train of waves it experiences an average steady resistance greater than that at the same speed in smooth water. There are no doubt several factors operative in producing this result ; one, for instance, may be described as wave pressure due to the reflexion or scattering of the ocean waves by the surface of the ship (Kreitner, 1939). This must certainly be taken into account in a complete theory, but investigation of it involves second-order terms in the hydrodynamical equations and a satisfactory solution of the problem would be difficult. Certain calculations which I have given recently (1940) seem to show that this cause is not likely to account for more than a small fraction of the observed results. Experiment shows that the effect is most prominent when the period of encounter of the ship with the waves is near the natural periods of the ship's oscillations ; whether directly or indirectly, the phenomenon is clearly associated with the heaving and pitching motions of the ship. In the paper already quoted (1940) it was suggested that it may well be that interactions between first-order effects which in themselves are purely periodic may, through phase differences, give rise to steady additional resistances. The object of the present note is to give some tentative calculations amplifying and illustrating this suggestion. For this purpose we fall back on the approximate theory which neglects the disturbing effect of the ship's surface upon the wave motion. In suitable cases we may perhaps regard the necessary additions for the reflected waves to be small corrections, as, for instance, for a long narrow ship (1937). This assumption was the basis of the theory developed by W. Froude in his work on the rolling of a ship among waves, in which case the wavelength is assumed large compared with the beam of the ship. It was also used explicitly by Kriloff in his well-known analysis of the heaving and pitching of a ship among waves. This latter work dealt only with the oscillations of the ship, and not with the extra resistance to motion which is now under consideration. It is true that the problem involves, in some form at least, second-order terms, and any partial separate examination of such terms is unsatisfactory ; the following calculations

---

* Communicated by the Author.

are therefore subject to correction by a more complete analysis, but they may serve to bring out a new point of view.

2. It is interesting to recall the development of the similar problem in rolling. Some years ago Suyehiro (1924), experimenting with a small model, announced the discovery of a drifting force sideways upon a ship when rolling in waves. This interesting result does not seem to have been studied by other workers, and either confirmed or disproved. The effect is small and probably needs suitable conditions of forced rolling in resonance with the natural period of roll. Suyehiro himself ascribed the force to reflexion of the waves by the side of the ship and supported this view by observation of the motion of the fluid particles near the ship. No calculation was made of the reflexion or scattering of the waves by the ship, and this is a problem which still awaits solution. Here, again, no doubt this form of wave pressure contributes to the result, but there is no reason to suppose it adequate in itself; moreover, the experiments showed a close association of the drifting force with the rolling of the ship. Recently an alternative theory has been put forward by Watanabé (1938). Starting from the Kriloff equations, Watanabé deduced an expression for the drifting force involving the angle of roll and the phase lag between the roll and the actuating moment; applied to Suyehiro's model, this expression gave a force of rather more than half the observed value.

In the following sections we derive similar expressions for the drifting force due to heaving and pitching when the ship is head-on to the waves; we assume throughout the usual theory of irrotational waves of small height.

3. Take the origin O in the undisturbed surface of the water, Ox horizontal and perpendicular to the wave crests and in the direction of the ship from stern to bow, Oy horizontal and Oz vertically upwards. We shall suppose first that the ship has no forward motion or, more precisely, we may suppose it constrained so that it is free to make small vertical oscillations and free also to make small rotational oscillations about a horizontal axis parallel to Oy through some point G. We consider plane waves of small amplitude h and of wave-length $2\pi/\kappa$ moving in the negative direction of Ox. To the first order the velocity potential is given by

$$\phi=(gh/\sigma)e^{\kappa z}\sin(\sigma t+\kappa x), \quad \ldots \quad (1)$$

with $\sigma^2=g\kappa$, and the pressure by

$$p=p_0-g\rho z+\rho\frac{\partial\phi}{\partial t} \quad \ldots \quad (2)$$

$$=p_0-g\rho z+g\rho h e^{\kappa z}\cos(\sigma t+\kappa x), \quad \ldots \quad (3)$$

p_0 being the pressure at the free surface.

A complete solution would include an addition to (1) necessary to satisfy the boundary condition at the surface of the ship in its actual motion and also the condition of constant pressure over the free surface of the water. We are, meantime, neglecting this additional term, and assuming the conditions such that for a first approximation we may calculate the resultant forces from the pressure given by (3). The resultant horizontal force backwards is given by

$$F=\iint pl\,dS, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

taken over the immersed surface of the ship in any position, (l, m, n) being the direction-cosines of the outward drawn normal at any point. This may be transformed into a volume integral taken throughout the immersed volume V of the ship, and using (3) we have

$$\begin{aligned} F&=\iiint \frac{\partial p}{\partial x}dV \\ &=-g\rho\kappa h\iiint e^{\kappa z}\sin(\sigma t+\kappa x)\,dV \quad . \quad . \quad . \quad . \quad . \quad (5)\end{aligned}$$

Let S_0, V_0 be the immersed surface and volume, respectively, when the ship is in its equilibrium position in smooth water. If the ship is held in this position in the waves, the corresponding force F_0 calculated from (5) is a purely periodic force with mean value zero (1937). Suppose now the ship to be in a slightly displaced position S due to heaving and pitching. The additional horizontal force is given by (5) integrated throughout the volume between S_0 and S. If $\delta\nu$ is the distance from any point of S_0 normally outwards to S, we have $dV=\delta\nu\,dS_0$. Let the pitch be measured by a small angle θ of rotation round an axis through a point G on Oz at a height c above O, taking θ to be positive with the bow up ; and let the heave be given by a small vertical displacement ζ upwards. Then, to the first order in ζ and θ, we have

$$\delta\nu=n\zeta+\{nx-l(z-c)\}\theta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Hence the horizontal force backwards in the new position is given by

$$\begin{aligned} F=F_0&-g\rho\kappa h\zeta\iint e^{\kappa z}\sin(\sigma t+\kappa x)n\,dS_0 \\ &-g\rho\kappa h\theta\iint e^{\kappa z}\sin(\sigma t+\kappa x)\{nx-l(z-c)\}\,dS_0, \quad . \quad (7)\end{aligned}$$

where the integrals are taken over the equilibrium position of the ship's surface.

Calculations may be made directly from this expression, but we put it into another form to show that it leads to an average steady force backwards.

Let B be the extra buoyancy for the ship in its equilibrium position due to the wave motion, that is, the additional force upwards arising from the term $\rho\partial\phi/\partial t$ in (2). Similarly, let P be the additional moment of this pressure about the axis through G in the direction of θ increasing. Then we have

$$B=-g\rho h\iint e^{\kappa z}\cos(\sigma t+\kappa x)n\,dS_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (8)$$

$$P=g\rho h\iint e^{\kappa z}\cos(\sigma t+\kappa x)\{l(z-c)-nx\}dS_0. \quad . \quad . \quad . \quad . \qquad (9)$$

Hence we may write (7) as

$$F=F_0-\frac{\kappa}{\sigma}\zeta\frac{\partial B}{\partial t}-\frac{\kappa}{\sigma}\theta\frac{\partial P}{\partial t}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (10)$$

The usual approximate equations for the motion of the ship are obtained by taking into account also the hydrostatic buoyancy and moment arising from the term $g\rho z$ in (3). With M, I as effective mass and moment of inertia, respectively, and assuming a simple law of damping in each case, the equations are

$$M\ddot{\zeta}+N\dot{\zeta}+g\rho A\zeta=B, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (11)$$

$$I\ddot{\theta}+N'\dot{\theta}+g\rho mV_0\theta=P, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (12)$$

A being the area of the water plane section and m the metacentric height for pitching.

When calculated from (8) and (9), B and P are of the form

$$B=B_0\sin(\sigma t+\alpha)\,;\quad P=P_0\sin(\sigma t+\alpha'), \quad . \quad . \quad . \quad . \qquad (13)$$

B_0, P_0, α, α' depending upon the wave-length and the form of the ship.

To obtain from (10) quadratic terms giving a mean value different from zero, we need consider only the forced oscillations in ζ and θ. These are given by

$$\left.\begin{array}{l}\zeta=kB_0\sin(\sigma t+\alpha-\beta)\\ \theta=k'P_0\sin(\sigma t+\alpha'-\beta')\end{array}\right\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (14)$$

k, k' being the usual magnification factors, and β, β' the corresponding phase lags, obtained by solving (11) and (12) for the forced oscillations. Using (13) and (14) in (10) and taking mean values of the quadratic terms, we obtain for the mean backward force

$$R=\tfrac{1}{2}\kappa kB_0{}^2\sin\beta+\tfrac{1}{2}\kappa k'P_0{}^2\sin\beta', \quad . \quad . \quad . \quad . \quad . \quad . \qquad (15)$$

an expression which is essentially positive.

With ζ_0 and θ_0 the amplitude of the forced heaving and pitching, respectively, and B_0, P_0 the amplitudes of the buoyancy and pitching moment as in (13), we have from (15)

$$R=\tfrac{1}{2}\kappa B_0\zeta_0\sin\beta+\tfrac{1}{2}\kappa P_0\theta_0\sin\beta'. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (16)$$

4. We have only used equations (11) and (12) to show that the average steady force is a resistance. In attempting comparison with experimental results one cannot rely upon calculations from these equations,

except for general descriptive purposes. Among other reasons, there is a lack of precise information about the damping of natural heaving and pitching. A common statement is that the damping in both cases is large, the natural oscillations in uniform waves diminishing rapidly and the motion reducing to the forced oscillations. On the other hand, this is difficult to reconcile with certain experimental results when the period of encounter with the waves is near a natural period; in such cases the amplitude of the resultant oscillation has a slow periodic variation from a minimum to a maximum in a manner suggesting the superposition of natural and forced oscillations of nearly equal period. The only published estimate from experimental results appears to be that given by Horn (1936), who states that the damping of heaving and pitching is of the same order of magnitude; his estimate gives a logarithmic decrement of about 1·4, an extremely large value compared, for instance, with the damping of rolling.

. We have assumed the ship to have no forward motion, but, so far as the present approximation goes, we may suppose it moving with uniform speed; the only difference is that the quantity σ in (13) is such that $2\pi/\sigma$ is the period of encounter of the ship with the waves.

We may make a rough estimate of the order of magnitude of the extra resistance given by (16). For a cargo boat of 400 ft. in waves of 500 ft. in wave-length and of height 6 ft., the amplitude P_0 of the pitching moment may be about 80,000 ft.-tons while the amplitude B_0 of the buoyancy might be, say, 300 tons. Hence from (16) we should have

$$R=1{\cdot}9\zeta_0 \sin\beta+502\,\theta_0 \sin\beta'. \qquad (17)$$

If the period of encounter is not near a natural period we might assume a total heave of 4 ft. and a pitch of 3°; whence

$$R=3{\cdot}8 \sin\beta+13 \sin\beta'. \qquad (18)$$

A value of 15° for the phase lags β, β' would not be unreasonable. This would give $R=4{\cdot}4$ tons, of which three-quarters would be associated with pitching and one-quarter with heaving.

5. For a more detailed analysis, we consider a simple form of wall-sided model of uniform draft d, with a rectangular middle body of length $2a$ and beam $2b$, and with entrance and run each of length l and of parabolic form. Thus with O at the centre of the parallel middle body, the equation of the contour from $x=a$ to $x=l+a$ is

$$y=b\{1-(x-a)^2/l^2\} \qquad (19)$$

at all depths, from $z=-d$ to $z=0$; and there is a similar equation for the run at the stern from $x=-(l+a)$ to $x=-a$.

The model is symmetrical fore and aft, and for simplicity we assume G, the position of the axis of rotation, to be at O; thus we take $c=0$ in (9). The integrations in (8) and (9) over the sides and keel of this model are readily carried out, and we obtain eventually

$$\mathrm{B}=\frac{8g\rho bh}{\kappa p^2}e^{-q}\{\sin(p+p_1)-p\cos(p+p_1)-\sin p_1\}\cos\sigma t. \qquad (20)$$

$$\begin{aligned}\mathrm{P}=\frac{8g\rho bh}{\kappa^2 p^2}[\{(p^2+p_1-3)\sin(p+p_1)-(p_1+3p)\cos(p+p_1)\\ +3\sin p_1-p_1\cos p_1\}e^{-q}+(1-e^{-q}-qe^{-q})\{\sin(p+p_1)\\ -p\cos(p+p_1)-\sin p_1\}]\sin\sigma t,\end{aligned} \qquad (21)$$

where $p=\kappa l$, $p_1=\kappa a$, $q=\kappa d$.

In (21) under usual conditions, the second part is small compared with the part which is factored by e^{-q}; if this latter factor be also neglected, the expression is simply the conventional pitching moment obtained from the hydrostatic pressure due to the wave elevation integrated over the water-plane section of the ship. This form of model is not quite suitable for the approximations on which the calculations have been made, but there are no experimental results available for simple symmetrical models of small beam.

From experiments, carried out at the National Physical Laboratory, on a model of a single-screw cargo ship, Kent and Cutland (1941) have obtained some very interesting results. We give the relevant data for the ship: length 400 ft., beam 55 ft., draught 24 ft., and displacement 11,332 tons. The natural pitching and heaving periods are given as 6·2 and 7·42 sec. respectively. Measurements were made of pitch and heave, of resistance and of other quantities, under various conditions in waves of 175 ft., 350 ft., and 490 ft. In the shortest waves it is probable that a considerable part of the increased resistance arises from the reflexion of the waves by the ship. Applying an expression which I gave recently (1940) for this resistance, Kent and Cutland show that it accounts for rather more than half the measured resistance for a ship moored in the waves. That expression gave a limiting value for a ship of great draught held at rest, assuming total reflexion by the front portion of the ship. It seems reasonable, therefore, to suppose that the force arising from reflexion would be much smaller for a ship of usual form floating on the water, and especially so for the large wave-lengths. Moreover, in the short waves the pitch and heave are slight and are irregular in period; in the medium waves the periods are approximately equal to the period of encounter between ship and wave, but the amplitude changes from a minimum to a maximum in regular cycles: in the long waves the pitch and heave are approximately uniform. In the present calculations the amplitudes in (16) are those of the forced oscillations of heave and pitch ; hence we attempt only a comparison with the results of Kent and Cutland for the 490 ft. wave-length, when presumably the motions are purely forced vibrations. Another reason for limiting comparison to waves of length greater than the ship is that the expressions for the buoyancy and pitching moment are probably better approximations than for shorter waves. Without attempting any close approximation to the form of the ship, we shall simply use the expressions (20) and (21) with

$$a=140 \text{ ft.} ; \quad l=60 \text{ ft.} ; \quad d=20 \text{ ft.} ;$$

$$b=27{\cdot}5 \text{ ft.} ; \quad h=2{\cdot}5 \text{ ft.} ; \quad \lambda=2\pi/\kappa=490 \text{ ft.}, \quad . \quad . \quad . \quad (22)$$

these dimensions giving a ship of about the same displacement, and waves of 490 ft. in length and 5 ft. in height.

With these values (20) and (21) give

$$\mathrm{B}=345 \cos \sigma t ; \quad \mathrm{P}=-83880 \sin \sigma t, \quad . \quad . \quad . \quad . \quad (23)$$

in tons, and ft.-tons respectively.

The numerical factors in (23) are the values of B_0 and P_0 in (16). The amplitudes ζ_0 and θ_0 we shall take from the observed results, assuming that these refer to forced vibrations in the period of encounter. The remaining factors are the phase differences, and these are more uncertain. It may be noted that the effect which is under discussion arises from the damping and the phase lag produced thereby ; if there is no phase lag there is no force. On the simple theory expressed in equations (11) and (12) the phase factor, $\sin \beta$ or $\sin \beta'$, has its maximum value of unity at resonance and diminishes on either side of this period. the diminution being more rapid the less the damping. The importance of the phase of the ship's motion in relation to that of the waves from a practical point of view is well known, but there are not many precise measurements suitable for the present purpose. The problem was attacked in an early paper by Kent (1922), and the recent paper by Kent and Cutland (1941) gives further detailed observations. They give a diagram showing positions of wave crest and trough along the ship at maximum pitch with the bow down, and from this one should be able to deduce the value of β' for use in (16). However, it must be remembered that the model was not a simple symmetrical form, with the axis of pitch at the centre of the water-plane section ; in fact, the position of this axis probably varied during the motion. It is also clear that the motion is not adequately covered by the theory of equations (11) and (12). For the authors state : " In general, as the ship's pitching period was not isochronous owing to the changing resistance to pitch, successive pitches showed a periodic movement of the wave crest position, backwards and forwards along the hull." The diagram given in the paper shows the mean positions. From this diagram, it seems that we may assume there was no measureable phase lag for the ship moored, with zero speed, in 490 ft. waves. Hence, from (16), the corresponding mean force is negligible, and this agrees with the observations. For the same wavelength when the ship had a speed of 8 knots, a rough estimate of the phase lag from the diagram is about 12·5°, and we take that value for β' in (16). As the free periods of heave and pitch are nearly equal, and the damping probably of the same order, we take the same value for β in (16). From the measurements given in the paper for 5 ft. waves, we have $\zeta_0=2{\cdot}1$ ft., $\theta_0=1{\cdot}6°$. With these values in (16) we get the extra resistance for the ship in tons; expressing the result for the model, 16 ft. long, we obtain from (16) a mean resistance of 0·63 lb. The measured value was 0·37 lb.

It is not worth while pursuing these tentative calculations further at present, but at least it seems that one can obtain results of the right order of magnitude; in fact, the calculated results are generally too high, especially at the peak values under resonance conditions, but that might have been anticipated. On the theoretical side, the various limitations and assumptions have already been sufficiently indicated in the course of the work. On the experimental side, there is a lack of suitable data obtained under conditions sufficiently approximating to the simplifications which have to be made before any calculations are possible.

6. In the present work, with the ship head-on to the waves, heaving and pitching have been considered together; for if the argument is valid for one kind of displacement it should also apply to the other. Moreover, the natural periods of heave and pitch are usually nearly equal and so resonance effects for the forced vibrations overlap. Reference has been made to Watanabé's work on the drifting force when rolling, that is, when the ship is broadside-on to the waves; in that work, as in Suyehiro's, the effect of heaving was entirely neglected. The expression (16) given here for the drifting force in heaving and pitching may also be used for heaving and rolling when the waves are broadside on, with the quantities in (16) having the appropriate values for those conditions. However, the natural periods of heave and roll usually differ considerably, and therefore the resonance effects are separated. The data for Suyehiro's model are not sufficient for these calculations to be made, otherwise one might compare the drifting forces due to heaving and rolling; it would be of interest if experiments could be devised to test whether these separate effects are observable, and to have experiments made under conditions suitable for comparison with theory.

Summary.

The problem considered is the drifting force on a ship when head-on to a regular train of waves. A satisfactory theory would have to include, among other factors, the effect of reflexion of the waves by the surface of the ship; in the present note this is neglected in order to make tentative calculations from another point of view which associates the effect directly with the oscillations of the ship. A steady average drifting force is obtained depending upon the phase differences between the heaving and pitching motions and the periodic forces and couples due to the wave motion. An examination is made of experimental results and, although available data are not suitable for detailed comparison, it appears that the calculations give drifting forces of the right order of magnitude.

References.

T. H. Havelock, Proc. Roy. Soc. A, clxi. p. 299 (1937).
T. H. Havelock, Proc. Roy. Soc. A, clxxv. p. 409 (1940).
F. Horn, *Jahrb. Schiffbau. Ges.* xxxvii. p. 153 (1936).
J. L. Kent, Trans. Inst. Nav. Arch. lxiv. p. 63 (1922).
J. L. Kent and R. S. Cutland, Trans. Inst. Engrs. and Shipbrs. Scotland, lxxxiv. p. 212 (1941).
J. Kreitner, Trans. Inst. Nav. Arch. lxxxi. p. 203 (1939).
K. Suyehiro, Trans. Inst. Nav. Arch. lxvi. p. 60 (1924).
Y. Watanabé, Trans. Inst. Nav. Arch. lxxx. p. 408 (1938).

XLVIII. *Newtonian Mass, its Definition and Determination.*

By Glen H. Draper *.

[Received March 8, 1942.]

Recently several articles have appeared in which the philosophy of Newtonian dynamics has been considered. In each case the author has had hidden from himself the complete hypotheses of a Newtonian isolated system, because he has not considered as important the question " What is Truth ? " †

Ernst Mach, in Germany, has done most of the valuable work in this connexion, but he has merely prohibited Newton's equations from defining themselves, and started several of the abstract, meaningless articles on the subject. The climax in abstraction came in the articles by Dr. C. G. Pendse, published in the ' Philosophical Magazine,' ser. 7, vols. xxiv., xxvi. and xxix., where it is mathematically " proven " that " it is impossible for a Newtonian observer to determine uniquely and exactly the ratios of the masses of the members of a system—an isolated system—of particles (as well as the reactions between them) from a knowledge of the motions of the particles if the system has more than seven particles, in the case where the observer utilizes the accelerations of the particles at as many instants as he chooses."

It is not the purpose of this note to discuss the errors found in the several articles, such as Pendse's calling his several observers in his Mass II. article by the name " Newtonian observers." It is the purpose of this note merely to restate a few of the prerequisite hypotheses imposed by the phrase " isolated Newtonian system of '*n*' material bodies."

1. Space has directional properties so that a co-ordinate system is " fixed " theoretically, but the determination of which is not explicitly assumed.

2. There is a singular point within an isolated system of " *n* " bodies that can be defined by means of the bodies and their mutual distances, which retains a constant position in the fixed co-ordinate system—the centre of gravity of the system. This point cannot be determined

---

* Communicated by the Author.

† Glen H. Draper, ' What is Truth ? ' Ransdell Inc., 1941.

independently of a knowledge of the relative masses of the " n " bodies, but, theoretically, it is assumed to be known whenever the equations of absolute motions are written.

3. If the centre of gravity is adopted as the origin of the " fixed " co-ordinate system, then the following set of simultaneous equations is necessarily true :—

$$\frac{d^s}{dt^s}\sum_i^n (m_i x_i)=0, \quad s \geqslant 2,$$

similar in y_i and z_i, since each vector may be resolved into an unique set of components parallel to the fixed axes.

4. All forces—causes of accelerations on each and every body within the system, arises and acts toward the bodies, the law being expressed as follows :—

$$m_i \ddot{x}_i = \sum_j^n m_j m_i \phi_{ij}(\ .\ .\ .\ )\frac{x_i - x_j}{r_{ij}},$$

similar expressions for y_i and z_i.

ϕ_{ij} is a variable which satisfies the several equations.

5. " Law III.—To every action there is an equal and opposite reaction, or the mutual reaction of two bodies is equal and opposite." That is,

$$\phi_{ji}=\phi_{ij}.$$

6. Each body is free to six degrees of freedom in its motion.

7. Variations of configurations among the bodies may take place.

If these seven conditions are accepted as the assumed hypotheses of the phrase " isolated Newtonian system of ' n ' material bodies," then a complete knowledge of the relative motions of the bodies is sufficient to determine uniquely the relative masses of the system, its centre of gravity, the co-ordinate directions, and the law of force, if " n " is greater than 2.

Hypothesis 7 is the condition which determines, *a priori*, if the masses can be determined specifically from a complete knowledge of the observable motions. If there are but two bodies, then the configuration is of necessity a straight line segment and therefore the relative masses *remain indeterminate*, even though the law of force is known also. The centre of gravity is not observable and therefore must be implied. A moment's reflexion should convince the reader that the relative masses of the solar system are obtained from perturbations that are produced because there are more than two bodies. This is a direct contradiction of Dr. Pendse's conclusion on page 1020 of his Mass I. article, where he asserts that the masses of the bodies of the solar system were obtained from the assumption " that the sun and one of the planets form an 'isolated' system." He further asserts that "the procedure is permissible only because it is fortunately possible to ignore the interactions between the planets in comparison with the gravitation of the sun." The masses are determined most exactly at present by the relations of motions between various two-body systems with an assumed law, *because* the accelerations are not known.

In conclusion, then, it is impossible to determine the relative mass in an "isolated Newtonian system" if the configurations which the bodies may assume are limited to fixed similar triangles, and but a single system, but in all other "isolated Newtonian systems" the masses are precisely determined, and the law of force is deduced from the definite knowledge of the motions of the bodies, since equations 3 and 4 are independent to the same degree that the configurations at different instants are independent.

NOTICE TO SUBSCRIBERS TO THE PHILOSOPHICAL MAGAZINE.

Subscribers are advised that for the period of the War, one volume (comprising 12 monthly parts) will be issued instead of two volumes (each comprising 6 monthly parts). The Contents and Index will appear with the December part.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

XLIX. *Units and Dimensions* *.

By E. A. Guggenheim, M.A., Sc.D.‡

[Received March 24, 1942.]

1. *Preamble.*

There would be little point in referring to all the articles and books which I have read on the subject under discussion, and still less in referring to those which I have not read. The striking feature of the subject is that it has remained controversial for so long. I am convinced that the main reason for this is loose terminology, and my object is to indicate how the greater part of the controversial matter, if not all of it, can be eliminated by a carefully chosen terminology and notation.

One of the leading physicists of the nineteenth century is sometimes quoted as having said that one can multiply together only numbers and that the idea of multiplying one length by another length is nonsense. I have not troubled to check exactly what he said, since I consider it a matter of small importance. We are under no obligation to accept as valid today any statement, however eminent the author, dating from an epoch when Newton's ideas on the motion of matter were regarded as sacrosanct, while his ideas on the propagation of light were with equal conviction rejected out of hand as wrong.

I adopt the opposite attitude that we are entitled to multiply together any two entities, provided our definition of multiplication is self-consistent and obeys the associative and distributive laws. Whether this extended conception of multiplication is useful or not is another question. In point of fact mathematicians find it profitable to multiply by one another, not only ordinary numbers, but also complex numbers, vectors, tensors, matrices, Dirac's *q*-numbers and various other entities §. It is likewise perfectly legitimate to multiply together any two physical entities, such as a length and a force. If the reader naïvely asks: "What, then, is the product of a foot and a pound?", I reply a "foot-pound". In case he suggests that this reply is unsatisfactory, I would point out that when a quarter is multiplied by three the product is "three-quarters", and when π is multiplied by $\sqrt{3}$ the product is $\sqrt{3}\pi$, and no simpler answer is possible.

---

* This article was written by request as the outcome of a discussion held by the Physical Society on February 6, 1942.

‡ Communicated by the Author.

§ The multiplication may or may not obey the commutative law. In the multiplication of physical scalar quantities in classical theory the commutative law is obeyed.

I shall in this article consistently make use of the conception of multiplying or dividing physical quantities by one another. If any reader, in spite of what I have already written, disputes my right to do this, it would be profitless for him to read any further.

2. *Notation.*

I shall throughout use letters without subscript or superscript to denote physical quantities. For example, if v denotes the speed of a ship, it would be correct to write

$$v=6\text{ knots}=10\text{ feet per second}=200\text{ yards per minute},$$

and wrong to write $v=6$ or 10 or 200 or any other number.

I shall use the corresponding letter with a superscript † to denote an arbitrarily chosen standard physical quantity of the same nature. With this notation it would be correct to write for the same ship's speed

$$\begin{aligned} v&=6\,v^\dagger, &\qquad v^\dagger&=1\text{ knot};\\ v&=10\,v^\dagger, &\qquad v^\dagger&=1\text{ foot per second};\\ v&=200\,v^\dagger, &\qquad v^\dagger&=1\text{ yard per minute}.\end{aligned}$$

Finally, I shall use the corresponding letter with a subscript m to denote the ratio of the physical quantity of interest to the chosen standard of the same nature. With this notation

$$v=v_m v^\dagger \quad\text{or}\quad v_m=v/v^\dagger,$$

and for the speed of the same ship we may write

$$\begin{aligned} v_m&=6, &\qquad v^\dagger&=1\text{ knot};\\ v_m&=10, &\qquad v^\dagger&=1\text{ foot per second};\\ v_m&=200, &\qquad v^\dagger&=1\text{ yard per minute}.\end{aligned}$$

The arbitrarily chosen quantity $v^\dagger$ is called the *unit*, and the corresponding number v_m is called the *measure* of the physical quantity v.

It is of vital importance to the understanding of the subject to distinguish clearly between the physical quantity v and its measure v_m. Apparently the majority of physicists prefer to use the plain letter without subscript to denote the measure. I, however, feel that the advantage of using the plain letter for the physical quantity is appreciated by anyone who has accustomed himself to this notation. This notation is by no means new; it was taught me by Alfred Lodge *, and to my knowledge was used by the late Professor Haigh.

3. *Definitional Constants.*

The area A of a rectangle is related to its length l and its height h according to

$$\mathrm{A}=hl. \qquad (3.1)$$

---

* Compare: A. Lodge, 'Nature,' xxxviii. p. 281 (1888). I discovered this letter only after I had completed my article. I am proud to find that several of the views expressed in §§ 1, 2, 3 and 6 are, in fact, a repetition of what my old teacher wrote over 50 years ago.

A formula for the area of a triangle with length of base l and height h is

$$A=\tfrac{1}{2}hl, \qquad (3.2)$$

and one for the area A of an ellipse of length (major axis) l and height (minor axis) h is

$$A=\frac{\pi}{4}hl. \qquad (3.3)$$

In principle it would be possible to define area in a different way so that the numerical constant became unity instead of $\frac{1}{2}$ in the second formula or became unity instead of $\pi/4$ in the last formula. But once the definition of area is settled, the numerical constant in each formula is uniquely determined, and, whatever definition be chosen, the numerical constant can be unity in at most one of these three formulæ. It will be useful to refer to such numerical constants as *definitional constants*.

A more instructive example of a definitional constant is the $\frac{1}{2}$ in the formula

$$T=\tfrac{1}{2}mv^2, \qquad (3.4)$$

relating the kinetic energy T to mass m and the velocity v. It would be possible in principle to re-define energy so that this numerical constant became unity, but the factor 2 would then appear in place of unity, as the definitional constant relating potential energy U to force f and distance x

$$f=-\frac{\partial U}{\partial x}. \qquad (3.5)$$

4. *Orthodox and Rational Definitions.*

We have seen that the value of a definitional constant can be changed by altering the definitions of the relevant physical quantities. In most cases, however, the same definition is universally accepted by all physicists. The definitions are then conveniently referred to as *orthodox*. For example, the formulæ of the preceding section are all in accordance with orthodox definitions, and any alternative definitions leading to other values of the definitional constants would not be orthodox.

Occasionally, however, it may be noticed that the definitions in current use lead to formulæ with an inconveniently large number of definitional constants containing π as a factor. This is apt to occur as the result of some definitional constant having been assigned the value unity in some system with circular or spherical symmetry, when it might more conveniently have been assigned the value unity in some other system with planar symmetry. When this situation arises it may be considered worth while introducing a new alternative definition, so as to reduce the number of π factors in the definitional constants. Such a revision is called *rationalization*, and I shall refer to the new definition as the *rational definition*. It is important to notice that the definitions, not the units, are rationalized.

The best known example of rationalization is that of electromagnetic definitions, but a discussion of this example is better postponed until other aspects of electromagnetic definitions have been clarified. Another important example of rationalization is Dirac's introduction of the constant $\hbar$ instead of Planck's constant $\mathbf{h}$. The fundamental quantum relation between a co-ordinate q and its conjugate momentum p then becomes

$$qp-pq=i\hbar$$

instead of $$qp-pq=i\mathbf{h}/2\pi,$$

and a multitude of quantal relations becomes similarly simplified. The relation between $\mathbf{h}$ and $\hbar$ is

$$\mathbf{h}=2\pi\hbar,$$

and it is to be noted that both quantities are measured in the same units, *e.g.* in erg sec.

5. *Dimensions.*

In classical mathematical analysis the expressions " tend to infinity ", " sum to infinity " and a few other expressions containing the word " infinity " have well-defined meanings, but the word " infinity ", apart from these expressions, is undefined and never used. In a like manner I shall now define two expressions containing the word " dimensions ", but I shall not have any other use for the word.

A pure number will be said to *have no dimensions* or to be *dimensionless*. If the ratio of two physical quantities is a pure number, the two quantities will be said to *have the same dimensions*.

According to this terminology any physical quantity Q has the same dimensions as the standard quantity $Q^\dagger$ of the same nature. Again, in any equations between physical quantities the two sides of the equation must have the same dimensions. The subject called *Dimensional Analysis* consists of using this fact either to verify that some formula is a possible one or even to show that, apart from an undetermined numerical constant, it is the only possible one.

6. *Germane Units.*

Let us now return to the formula relating kinetic energy T to mass m and velocity v:

$$T=\tfrac{1}{2}mv^2. \qquad (6.1)$$

In our terminology the numerical factor on the right is a *definitional constant* and its value is $\frac{1}{2}$ according to the *orthodox definitions*.

If we express each physical quantity as the product of its measure and the standard chosen as unit, we obtain

$$T_m T^\dagger=\tfrac{1}{2}m_m v_m^{\,2} m^\dagger v^{\dagger 2}, \qquad (6.2)$$

an identity valid for any choice of the units $T^\dagger$, $m^\dagger$ and $v^\dagger$. Let us now choose our units in such a way that

$$T^\dagger=m^\dagger v^{\dagger 2}. \qquad (6.3)$$

Substituting from (6.3) into (6.2) we obtain

$$T_m=\tfrac{1}{2}m_m v_m^{\,2}, \qquad (6.4)$$

a relation of precisely the same form as the relation (6.1). It is important to appreciate the difference in meaning between these two relations of the same form. Whereas formula (6.1) is true, according to orthodox definitions, whatever units be used, the validity of (6.4) is entirely dependent on a choice of units such that the relation (6.3) is satisfied.

When a set of units conforms to relations like (6.3), I will call it a set of *germane units.* I am deliberately avoiding the epithet " consistent ", because as far as formula (6.1) is concerned any units are consistent. The units need to be germane only if we require formula (6.1) to reduce to the formula (6.4).

The following are three examples of sets of germane units.

(a) $m^\dagger = \text{lb.}$, $v^\dagger = \text{ft./sec.}$, $T^\dagger = \text{lb.ft.}^2/\text{sec.}^2 = \text{ft.poundal}$;

(b) $m^\dagger = \text{g.}$, $v^\dagger = \text{cm./sec.}$, $T^\dagger = \text{g.cm.}^2/\text{sec.}^2 = \text{erg}$;

(c) $m^\dagger = \text{kg.}$, $v^\dagger = \text{m./sec.}$, $T^\dagger = \text{kg.m.}^2/\text{sec.}^2 = \text{joule}$.

7. *Numerical Ratios.*

If v denotes the speed of light in some medium and $\mathbf{c}$ its speed in empty space, then v and $\mathbf{c}$ have the same dimensions. The ratio $\mathbf{c}/v$ is therefore a pure number which we denote by n and call the *refractive index* of the medium. For example, we may write

$$\mathbf{c} = 3 \times 10^8 \text{ m./sec.} = 1 \text{ light-year/year},$$

$$v = 2 \times 10^8 \text{ m./sec.} = \tfrac{2}{3} \text{ light-year/year},$$

$$n = 3/2.$$

With the second choice of units, namely, $v^\dagger = \text{light-year/year}$, the measure v_m of v is equal to the reciprocal of the refractive index. We may therefore write

$$v_m n = 1, \qquad v^\dagger = \text{light-year/year}.$$

The point to be emphasized is that the refractive index n is, according to definition, a number, and by particular choice of units the measure v_m of v can be made equal to the reciprocal of the refractive index, but v itself is a speed and cannot be equal to a number.

The above remarks may seem trivial, and indeed, as long as we are careful with our terminology, they are trivial. If, however, we were careless enough to confuse a light-year with a year, we should fall into confusion comparable to that which has raged in electrostatics and has been entirely due to loose terminology.

This matter of distinguishing between a physical quantity on the one hand and the ratio of this physical quantity to another one of the same nature (and so of the same dimensions) on the other is so important that it seems worth while calling attention to further examples before proceeding to a discussion of electric quantities.

If we define *density* as mass per unit volume, then the density of water is 1 kg./l. or $62\frac{1}{2}$ lb./ft.^{3} It is obvious that density is not a pure number.

On the other hand, *specific gravity*, defined as the ratio of the density of a substance to that of water, is by definition a pure number. For example, mercury has a density 13·6 kg./l. = 13·6 g./cm.^{3} = 850 lb./ft.^{3} and a specific gravity 13·6. Here there is no confusion because there are different well-established names for *density* and *specific gravity*.

In my next example the terminology is not universal and consequently there is sometimes confusion. If we define *heat capacity* as the energy required (under certain specified conditions) to produce unit rise of temperature in unit quantity of a substance, then the heat capacity of water is (in round figures) 1 B.T.U./lb. deg. F. or 4·2 joules/g. deg. C. or 18 cal./mole deg. C. If, further, we define *specific heat* as the ratio of the heat capacity of a substance to that of water (under specified standard conditions), then specific heat is evidently a number by definition. In this case it is obvious from my definitions that heat capacity is not a pure number and that specific heat is one, but these definitions are not universal. Whenever an author is careless enough to use the same name for what I have called *heat capacity* and what I have called *specific heat*, it is impossible to predict the limits of the confusion into which he may fall.

Incidentally, this seems an appropriate place to protest against the statement sometimes made that the entropy of a mole of substance is so many "entropy units", a term which might equally well mean B.T.U./deg. F. as joules/deg. C. The statement is, therefore, just as informative as saying that the pressure of a gas is so many pressure units.

8. *The Franklin.*

A few months ago I suggested * that the electric charge which in empty space repels a similar charge at a distance of one centimetre with a force of one dyne be called a *franklin*, in honour of Benjamin Franklin, the pioneer of static electricity. It is too early for me to judge whether this suggestion will find favour. Since, however, a name for this charge is required for the ensuing discussion, I shall use the name *franklin*.

9. *Permittivity and Dielectric Constant.*

If f denotes the force of attraction between two stationary charges q_A and q_B in empty space at a distance apart r, then, according to Coulomb's law,

$$-\frac{q_A q_B}{f r^2} = \mathbf{a}, \qquad \qquad (9.1)$$

where **a** is a universal constant. Its value is given (in round numbers) by

$$\mathbf{a} = 1\,\frac{\text{franklin}^2}{\text{erg cm.}} = \frac{1}{9\times10^9}\,\frac{\text{coulomb}^2}{\text{joule m.}} = \frac{1}{9\times10^9}\,\frac{\text{sec.}}{\text{ohm m.}}. \qquad (9.2)$$

---

* Guggenheim, 'Nature,' cxlviii. p. 751 (1941).

If we now consider the force between the same pair of charges the same distance apart in a medium, we may write, in place of (9.1),

$$-\frac{q_A q_B}{fr^2}=\epsilon \mathbf{a}, \qquad \ldots \ldots \quad (9.3)$$

where ϵ is characteristic of the medium. It is obvious by comparison of (9.3) with (9.1) that $\epsilon\mathbf{a}$ has the same dimensions as **a**, and so ϵ is a pure number.

Thus described the situation is free from ambiguity, which can, however, arise through loose terminology if the same name is used for $\epsilon\mathbf{a}$ and ϵ. All confusion is avoided if $\epsilon\mathbf{a}$ is called the *permittivity of the medium*, **a** the *permittivity of empty space*, and ϵ the *dielectric constant.*

A comparison of the present section with § 7 will be found instructive. The dielectric constant of empty space is unity by definition, as are the refractive index of empty space, the specific gravity of water, and the specific heat of water. On the other hand, the permittivity **a** of empty space has a value given by (9.2), and so the statement

$$\mathbf{a}=1\,\frac{\text{franklin}^2}{\text{erg cm.}}$$

is analogous to the statements

$$\text{speed of light in empty space}=1\,\frac{\text{light-year}}{\text{year}};$$

$$\text{density of water}=1\,\frac{\text{kg.}}{\text{l.}};$$

$$\text{heat capacity of water}=1\,\frac{\text{B.T.U.}}{\text{lb. deg. F.}}.$$

In all four statements the numerical coefficient is made unity by suitable preselection of the units used.

10. *Sets of Germane Units for Electrostatic Quantities.*

A physicist who properly understands units feels equally happy with any set of germane units and has an infinite variety of choice to suit his convenience in any particular problem. It is unlikely to be convenient to use the same units in a cosmological problem as in a subatomic one. There is, however, a certain school who wish to impose an obligation on all physicists and engineers always to use the same set of units. In my opinion this attitude is virtually a confession of not having sufficient grasp of units to be able to switch readily from one set to another. It should be equally easy to calculate with feet as with metres, with hours as with seconds, with joules as with ergs, with deg. F. as with deg. C., with franklins as with coulombs. Difficulty arises only when there is no adequate name for the units one wants to use. The introduction of the name *franklin* dispels all difficulties over the e.s.u., which I would prefer to call the c.g.s.f. (centimetre-gram-second-franklin) system. A comparison between this system and the m.k.s.c. (metre-kilogram-second-coulomb) system is summarized in Table I.

The reader should have no difficulty in constructing for himself another set of germane units based on the centimetre-gram-second-dekacoulomb. This is, in fact, the system generally known as the e.m.u. system. Since, so far as I know, this system is used only in elementary courses and old-fashioned text-books, there does not seem to be any need for a shorter name for the dekacoulomb. In these units the permittivity of empty space is

$$\mathbf{a}=\frac{1}{9\times10^{20}}\,\frac{\text{dekacoulomb}^2}{\text{erg cm.}}.$$

TABLE I.

Term.	c.g.s.f. unit (e.s.u.).	m.k.s.c. unit.
Charge	franklin	coulomb
Surface density of charge	franklin/cm.^{2}	coulomb/m.^{2}
Displacement, D	franklin/cm.^{2}	coulomb/m.^{2}
Current	franklin/sec.	coulomb/sec. = amp.
Current density	franklin/sec. cm.^{2}	amp./m.^{2}
Energy	erg	joule
Energy density	erg/cm.^{3}	joule/m.^{3}
Potential	erg/franklin	joule/coulomb = volt
Field, E	dyne/franklin	volt/m.
Permittivity, D/E	franklin^2/erg cm.	coulomb/volt m. = sec./ohm m.
Value of permittivity of empty space	$1\,\dfrac{\text{franklin}^2}{\text{erg cm.}}$	$\dfrac{1}{9\times10^9}\,\dfrac{\text{sec.}}{\text{ohm m.}}$

11. *Rationalization of Electrostatic Relations.*

In §4 the example given of rationalization was Dirac's introduction of the universal constant $\hbar$ to take the place of $\mathbf{h}$. Rationalization in electrostatics is somewhat similar and consists essentially in the introduction of a universal constant $\boldsymbol{a}$ related to $\mathbf{a}$ by

$$\boldsymbol{a}=\mathbf{a}/4\pi. \qquad (11.1)$$

For the force of attraction f between stationary electric charges q_A and q_B a distance r apart in a medium of dielectric constant ϵ, we may write, as an alternative to (9.3),

$$f=-\frac{q_Aq_B}{4\pi\epsilon\boldsymbol{a}r^2}. \qquad (11.2)$$

At the same time in the rational system the electric displacement $\mathscr{D}$ is related to the ordinary displacement D by

$$\mathscr{D}=4\pi\mathrm{D}. \qquad (11.3)$$

The simplification of formulæ due to rationalization can be illustrated by considering a parallel plate condenser, neglecting the edge effect. If the area of the plates is A, their separation is l, the charges on the plates are $\pm q$ and the dielectric constant of the medium is ϵ, then the formulæ for the various derived quantities are collected in Table II. according to both the rational system and the ordinary (irrational) system.

It is clear that the effect of rationalization is the disappearance of numerous factors 4π in the formulæ applying to a parallel plate condenser. There will be a corresponding increase in the π factors in formulæ applying to systems with spherical symmetry. For example, the potential V at a distance r from a single point charge q in a medium of dielectric constant ϵ is given by

$$V = \frac{q}{\epsilon \mathbf{a} r} = \frac{q}{4\pi\epsilon \boldsymbol{a} r} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11.4)$$

It is to be emphasized that all the above formulæ, whether those of the rational system or of the ordinary system, are entirely independent of any choice of units. In any of these formula the unit of charge may be the coulomb or the franklin or any other unit. The value of **a** in the c.g.s.f. and the m.k.s.c. systems is given by (9.2) and the corresponding values of ***a*** are derived immediately by means of (11.1).

TABLE II.

	Rational definitions.	Ordinary definitions.
Total normal induction between plates	q	$4\pi q$
Electric displacement normal to plates	$\mathscr{D} = q/A$	$D = 4\pi q/A$
Field intensity	$E = q/\epsilon \boldsymbol{a} A$	$E = 4\pi q/\epsilon \mathbf{a} A$
Force between plates	$q^2/\epsilon \boldsymbol{a} A$	$4\pi q^2/\epsilon \mathbf{a} A$
Potential difference between plates.	$ql/\epsilon \boldsymbol{a} A$	$4\pi ql/\epsilon \mathbf{a} A$
Electrostatic energy	$\frac{1}{2} q^2 l/\epsilon \boldsymbol{a} A$	$2\pi q^2 l/\epsilon \mathbf{a} A$
Energy density	$\frac{1}{2} q^2/\epsilon \boldsymbol{a} A^2 = \frac{1}{2} E \mathscr{D}$	$2\pi q^2/\epsilon \mathbf{a} A^2 = \frac{1}{2} ED/4\pi$
Capacity	$\epsilon \boldsymbol{a} A/l$	$\epsilon \mathbf{a} A/4\pi l$

12. *Permeability and Magnetic Constant.*

If two charges q_A and q_B a distance r apart are moving parallel to each other with a velocity v at right angles to the line joining them, then the force of attraction f between them in empty space is of the form

$$f = -\frac{q_A q_B}{\mathbf{a} r^2}\left(1 - \frac{v^2}{\mathbf{c}^2}\right), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12.1)$$

where **c** is a universal constant, which must obviously have the same dimensions as v. From the theory it turns out that the speed of propagation of electromagnetic effects is equal to **c**.

An equivalent statement to (12.1) is that, apart from the electrostatic repulsion between electric charges, there is a magnetic attraction between current elements ; in particular, the magnetic attractive force f between two current elements $i_A ds_A$ and $i_B ds_B$ parallel to each other and at right angles to the line of length r joining them is in empty space given by

$$f=\frac{i_A ds_A i_B ds_B}{\mathbf{a}\mathbf{c}^2 r^2} \qquad (12.2)$$

The corresponding force between the same two current elements similarly situated in a medium is given by

$$f=\frac{\mu i_A ds_A i_B ds_B}{\mathbf{a}\mathbf{c}^2 r^2}, \qquad (12.3)$$

where μ is characteristic of the medium. It is obvious by comparison of (12.3) with (12.2) that μ is a pure number.

Thus described the situation is free from ambiguity, which will, however, arise if the same name is used for the pure number μ and the physical quantity $\mu/\mathbf{ac}^2$. The universal constant $1/\mathbf{ac}^2$ occurring in (12.2) has the value

$$\frac{1}{\mathbf{ac}^2}=\frac{1}{9\times10^{20}}\frac{\text{g. cm.}}{\text{franklin}^2}=\frac{1}{10^7}\frac{\text{kg. m.}}{\text{coulomb}^2}$$
$$=\frac{1}{10^7}\frac{\text{ohm sec.}}{\text{m.}}=1\frac{\text{g. cm.}}{\text{dekacoulomb}^2}. \qquad (12.4)$$

and is called the *permeability of empty space*. Consistency practically compels us to call the constant $\mu/\mathbf{ac}^2$ in (12.3) the *permeability of the medium*. We still require a name for μ, and in order to avoid confusion it is important that μ should not be called permeability. I suggest the following new terminology :—

The ratio μ of the permeability of a medium to the permeability of empty space should be called the *magnetic constant of the medium*.

With this terminology the *magnetic constant* μ, like the dielectric constant ϵ, is a pure number, and these two numbers are related to the *refractive index*, also a pure number, by

$$\epsilon\mu=n^2. \qquad (12.5)$$

The product of the permittivity $\epsilon\mathbf{a}$ and the permeability $\mu/\mathbf{ac}^2$ is thus equal to $n^2/\mathbf{c}^2$, that is to the reciprocal of the square of the speed of light in the medium.

The terminology here proposed would entail only one small alteration in the tabulation of experimental data. Tables of μ would have to be renamed tables of *magnetic constants*. Fortunately even this small change would be required but rarely, because the quantity most usually tabulated is not μ, but κ, defined by

$$\kappa=\frac{\mu-1}{4\pi}, \qquad (12.6)$$

and there is no objection to continuing to call κ the *susceptibility*.

13. *Electromagnetic Field Quantities.*

All observable electromagnetic phenomena can in principle be described in terms of interactions between electric charges and currents. It is, however, convenient for the mathematical development of the subject to introduce the four field vectors called the electric displacement D, the electric field strength E, the magnetic field strength H and the magnetic induction B.

The electric displacement D and the magnetic field strength H are in a stationary state completely determined by the distribution of electric charges and currents respectively. The electric field strength E and the magnetic induction B are measures of the force on unit charge and current element respectively. In this respect H is the magnetic analogue of D, while B is the magnetic analogue of E. The nomenclature is lamentable, but presumably too long established to be changed.

We have already noted in §10 that D has the same dimensions as charge/area and E has the same dimensions as force/charge, so that the product ED has the same dimensions as pressure. The dimensions of H are the same as those of current/distance and those of B are those of force/(current × distance), so that the product BH has the same dimensions as pressure. A full appreciation of the symmetry in the correlation between D, E on the one hand and H, B on the other can best be grasped in terms of special relativity. In this connexion I would refer the reader to an extremely illuminating article by Sommerfeld * entitled " Uber die Dimensionen der Elektromagnetischen Grössen ". It would take up too much space to refer to this aspect of the subject in any detail and we must here content ourselves with a simple verification of the above statements concerning the dimensions of H and B.

If a straight long conductor carrying a current i is surrounded by a uniform medium of magnetic constant μ, then at a distance r from the conductor H and B are given by

$$\mathrm{H}=2\frac{i}{r}, \quad \mathrm{B}=2\frac{\mu}{\mathbf{a}\mathbf{c}^2}\frac{i}{r}. \qquad \ldots\ldots \quad (13.1)$$

If, instead of a long straight conductor, we consider a circular one of radius r, then at the centre of the circle

$$\mathrm{H}=\frac{i}{r}, \quad \mathrm{B}=\frac{\mu}{\mathbf{a}\mathbf{c}^2}\frac{i}{r}. \qquad \ldots\ldots \quad (13.2)$$

Again, if we consider a solenoid with one turn per length l, then inside the solenoid

$$\mathrm{H}=4\pi\frac{i}{l}, \quad \mathrm{B}=4\pi\frac{\mu}{\mathbf{a}\mathbf{c}^2}\frac{i}{l}. \qquad \ldots\ldots \quad (13.3)$$

The numerical factors 2, 1 and 4π in these formulæ are *definitional constants* and the values given correspond to the ordinary definitions. Rationalization would lead to different numerical values of the definitional constants, without in any way affecting the dimensional relations. The first formula of (13.1) or (13.2) or (13.3) verifies that H has the same dimensions as current/distance. Multiplication of any first formula by the second formula and comparison with (12.3) confirms that the product BH has the same dimensions as $f/ds_A\,ds_B$, that is to say, as force/area or pressure. Consequently B has the same dimensions as force/(current × distance).

---

* Sommerfeld, *Zeit. techn. Physik*, xi. p. 420 (1935).

In the m.k.s.c. system the unit of B is

$$B^{\dagger} = 1\,\frac{\text{joule}}{\text{m.}^2\,\text{amp.}} = 1\,\frac{\text{volt sec.}}{\text{m.}^2} = 1\,\frac{\text{weber}}{\text{m.}^2}, \quad \ldots \quad (13.4)$$

and that of H is

$$H^{\dagger} = 1\,\frac{\text{amp.}}{\text{m.}} = 1 \text{ millioersted.} \quad \ldots \quad (13.5)$$

If the m.k.s.c. system is to be widely used, I hope some new and shorter name will be introduced for the millioersted, but I cannot think of an obviously suitable one.

The permeability of empty space in the m.k.s.c. system is

$$\frac{1}{\mathbf{a}\mathbf{c}^2} = \frac{1}{10^7}\,\frac{\text{weber}}{\text{m.}^2\text{ millioersted}} = \frac{1}{10^7}\,\frac{\text{volt sec.}}{\text{amp. m.}} = \frac{1}{10^7}\,\frac{\text{ohm}}{\text{m.}}\,\text{sec.}$$
$$= \frac{1}{10^7}\,\frac{\text{joule}}{\text{amp.}^2\,\text{m.}} = \frac{1}{10^7}\,\frac{\text{kg. m.}}{\text{coulomb}^2}. \quad (13.6)$$

14. *Pipe and Crack Fields.*

Owing to the fact that any measuring instrument occupies space it is not possible to measure directly any of the field vectors D, E, H, B inside a medium, and their values have to be computed from measured values in suitably shaped cavities. It is of particular interest to consider a narrow pipe-shaped cavity parallel to the electric or the magnetic field and a narrow crack-shaped cavity across the field. Following a descriptive nomenclature proposed by Darwin *, I shall refer to the fields in these two kinds of cavity as the *pipe field* and the *crack field* respectively. I shall further denote the pipe fields by using the subscript p and the crack fields by using the subscript c. Using the fact that at an interface there is continuity of the tangential components of E, H and of the normal components of D, B, we have, as relations between the internal field and the corresponding pipe and crack fields,

$$\mathbf{a}E_c = D_c = D = \epsilon\mathbf{a}E = \epsilon\mathbf{a}E_p = \epsilon D_p, \quad \ldots \quad (14.1)$$

$$\frac{H_c}{\mathbf{a}\mathbf{c}^2} = B_c = B = \frac{\mu H}{\mathbf{a}\mathbf{c}^2} = \frac{\mu H_p}{\mathbf{a}\mathbf{c}^2} = \mu B_p. \quad \ldots \quad (14.2)$$

These relations are independent of any choice of units and with the notation used there is no possible ambiguity. Confusion arises only if the same symbol is used for two or more different quantities.

15. *Rationalization of Electromagnetic Relations.*

The rationalization of electrostatic quantities has already been described in §11. The extension of the process to electromagnetic quantities follows similar lines. I shall not give details, but will merely illustrate the process by a comparison between Maxwell's equations in the ordinary (irrational) and in the rational form. Using $\boldsymbol{a}$, $\mathscr{D}$, $\mathscr{H}$ for the rational quantities corresponding to the ordinary (irrational) $\mathbf{a}$, D, H, and denoting current density by I, we have the following equations :—

---

* Darwin, Proc. Roy. Soc. A, cxlvi. p. 17 (1934).

$$\begin{array}{lll}
\textit{Ordinary (irrational)} & \textit{Rational} & \\
\text{curl } \mathrm{H} = 4\pi \mathrm{I} + \frac{d\mathrm{D}}{dt}, & \text{curl } \mathscr{H} = \mathrm{I} + \frac{d\mathscr{D}}{dt}, & \quad (15.1) \\
\text{curl } \mathrm{E} = -\frac{d\mathrm{B}}{dt}, & \text{curl } \mathrm{E} = -\frac{d\mathrm{B}}{dt}, & \quad (15.2) \\
\text{div } \mathrm{D} = 4\pi\rho, & \text{div } \mathscr{D} = \rho, & \quad (15.3) \\
\text{div } \mathrm{B} = 0, & \text{div } \mathrm{B} = 0, & \quad (15.4) \\
\mathrm{D} = \epsilon \mathbf{a} \mathrm{E}, & \mathscr{D} = \epsilon \boldsymbol{a} \mathrm{E}, & \quad (15.5) \\
\mathrm{B} = \frac{\mu \mathrm{H}}{\mathbf{a}\mathbf{c}^2}, & \mathrm{B} = \frac{\mu \mathscr{H}}{\boldsymbol{a}\mathbf{c}^2}. & \quad (15.6)
\end{array}$$

For a non-conducting medium I=0, and substitution from (15.5) into (15.1) and from (15.6) into (15.2), followed by elimination of H or $\mathscr{H}$ leads to the same equation in both the ordinary and the rational system, namely

$$\nabla^2 \mathrm{E} = \frac{\epsilon\mu}{\mathbf{c}^2} \frac{d^2\mathrm{E}}{dt^2}. \qquad (15.7)$$

From this it is seen that the speed of wave propagation is $\mathbf{c}/\sqrt{\epsilon\mu}$ and that the refractive index is equal to $\sqrt{\epsilon\mu}$, as already mentioned in § 12.

Whether rationalization be used or not, I think it will be admitted that Maxwell's equations (15.1) to (15.4) appear simpler in the present notation, where the symbols denote physical quantities, than in any alternative notation, where the symbols denote the measures referred to special sets of units.

16. *Derived Universal Constants.*

Our ability to write down physical formulæ independent of any choice of units is the result of using simple symbols to denote physical quantities, not their measures. It is, therefore, important not to omit from any formula the symbol for any physical quantity merely on the grounds that in some particular unit its measure is one. In particular, the permittivity of empty space $\mathbf{a}$ must not be omitted on the grounds that in the c.g.s.f. system its value is 1 franklin²/erg cm. Consequently $\mathbf{a}$ will occur as a factor in many expressions from which it has hitherto been omitted on the tacit understanding that the expression was valid only in the e.s.u. (or c.g.s.f.) system of units. This applies to the expressions for several universal constants derived from other universal constants considered more fundamental.

Two examples will suffice. The fine structure constant $\mathbf{ahc}/2\pi\mathbf{e}^2$ is a pure number equal to 137 and Rydberg's constant $2\pi^2 m\mathbf{e}^4/\mathbf{a}^2\mathbf{ch}^3$, where

m is the reduced mass of the electron and the proton, is a reciprocal length. These statements are true whatever units are used.

17. *Numerical Examples.*

One of the advantages of using symbols, as we have done, to denote physical quantities instead of their measures is that when values are substituted for the symbols one obtains an automatic check on the necessary equality of dimensions of the two sides of any equation. This advantage is particularly striking when not all the units used are germane to one another. It seems worth while illustrating this point by an actual example.

Let us suppose that it is required to activate a solenoid of cross-sectional area A with a current i through n turns of wire of density ρ and resistivity σ. It can be shown by elementary reasoning that the product of the mass M of wire required and the power P used is given by

$$\mathrm{PM}=4\pi\rho\sigma\mathrm{A}(ni)^2. \qquad (17.1)$$

The interesting feature * of this formula is that it is independent of the thickness of the wire used. Since our symbols denote, as usual, physical quantities (not their measures), this formula is independent of any choice of units.

Suppose, now, we decide to use copper wire and elect to measure P in watts, M in lb., A in in.², and ni in amp. turns. We accordingly transcribe (17.1) to the form

$$\frac{\mathrm{P}}{\text{watt}}\,\frac{\mathrm{M}}{\text{lb.}}=k\frac{\mathrm{A}}{\text{in.}^2}\left(\frac{ni}{\text{amp.}}\right)^2, \qquad (17.2)$$

where k is a numerical constant defined by

$$k=4\pi\rho\sigma\,\frac{\text{in.}^2\,\text{amp.}^2}{\text{watt lb.}} \qquad (17.3)$$

For copper $\rho=555$ lb./ft.³ and $\sigma=1{\cdot}77\times10^{-6}$ ohm cm.; substituting these values into (17.3) we obtain

$$\begin{aligned}k&=4\pi\times555\times1{\cdot}77\times10^{-6}\,\frac{\text{lb. ohm cm. in.}^2\,\text{amp.}^2}{\text{ft.}^3\,\text{watt lb.}},\\&=4\pi\times555\times1{\cdot}77\times10^{-6}\,\frac{\text{cm.}}{\text{ft.}}\,\frac{\text{in.}^2}{\text{ft.}^2},\\&=\frac{4\pi\times555\times1{\cdot}77\times10^{-6}}{30{\cdot}5\times144}=2{\cdot}82\times10^{-6}. \qquad (17.4)\end{aligned}$$

I suggest that no alternative method of expressing the relation (17.2) or of evaluating the constant k is more illuminating or more fool-proof.

It is instructive to consider the application of our notation to formulæ involving logarithms or exponential functions. Let us take as an example the formula for the vapour pressure p of an inert element of the argon group. The theoretical formula obtained by statistical mechanics is

---

* I am indebted to Mr. S. Butterworth for drawing my attention to this relation.

$$\log_e p = -\frac{\chi}{\mathbf{k}T} - \iint + \log_e \frac{(2\pi m \mathbf{k}T)^{3/2}\mathbf{k}T}{\mathbf{h}^3}, \quad . \quad . \quad . \quad (17.5)$$

where m is the mass of a molecule (atom) and χ is the energy required to take an atom from the crystal at the absolute zero to a state of rest at infinite separation ; $\iint$ is an abbreviation for

$$\int_0^T \frac{dT'}{T'^2} \int_0^{T'} \frac{C_P}{\mathbf{k}} dT'', \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.6)$$

where C_P denotes the molecular heat capacity at constant pressure of the crystal. Since the left side of (17.5) is the logarithm of a pressure, whereas the first two terms on the right are pure numbers, the last term on the right must also be the logarithm of a pressure, and it can readily be verified that such is the case. For numerical application to the several substances of the argon group it is convenient to transcribe (17.5) to one of the forms

$$\log_e \frac{p}{\text{atm.}} = -\frac{\chi}{\mathbf{k}T} - \iint + \frac{5}{2}\log_e \frac{T}{\text{deg. C.}} + \frac{3}{2}\log_e M - 3{\cdot}655, \quad . \quad . \quad (17.7)$$

$$\log_{10} \frac{p}{\text{atm.}} = -\frac{0{\cdot}4343\chi}{\mathbf{k}T} - 0{\cdot}4343\iint + \frac{5}{2}\log_{10} \frac{T}{\text{deg. C.}} + \frac{3}{2}\log_{10} M - 1{\cdot}587, \quad . \quad . \quad . \quad (17.8)$$

where M is the chemical " molecular weight " defined as the ratio of the molecular mass to one sixteenth of the molecular mass of oxygen, and is, in spite of its name, a pure number. Written thus every term of (17.7) and (17.8) is a pure number, and there is no possibility of ambiguity or confusion over units.

Our notation (letters denoting physical quantities) leads to a neat and compact way of heading tables of data or labelling the co-ordinate scales of a graph. If, for example, the physical quantity tabulated or plotted is a pressure and the unit chosen the atmosphere, we should head the table or label the scale P/atm., whereas authors who use P to denote the numerical measure of the pressure have to write P (in atm.) or $P_{(\text{atm.})}$, or something similar. I consider that P/atm. is more expressive and less cumbrous than any of these alternatives.

18. *Number of Independent Dimensions.*

Up to this point I have tried to remove controversy, contradiction, and ambiguity by taking extra care over my notation and terminology. As far as this object is concerned, my article should end here and the views which I am now about to express may be controversial. They do not, however, affect anything in the previous sections.

Up to the present I have restricted the use of the word dimensions to statements such as the following :—the dimensions of velocity are the same as those of a length divided by a time. I shall now extend the use of the word dimensions to include statements such as :—the dimensions of velocity can be expressed in terms of the dimensions of length and time. We can now consider the question which has led to the greatest controversy. If we regard certain physical quantities as fundamental, what is the least number of fundamental quantities such that the dimensions of all physical quantities can be expressed in terms of the dimensions of the fundamental quantities ? It is to be noted that, once this number is settled, there is still an infinite variety of choice as to which physical quantities are taken as fundamental.

According to Rücker* the required number of fundamental quantities is five, and Tolman ‡ gives the reason that physics considers five fundamental different kinds of " thing ", namely space, time, matter, electricity and entropy. If we provisionally accept the number five as right, we may, for example, make our choice in the following ways amongst others :

Length, time, mass, charge, temperature (Most usual.)

Length, time, mass, charge, entropy. (Tolman's choice.)

Length, time, energy, charge, temperature. (My personal preference.)

Instead of considering the question *ab initio*, I propose to take the answer first given by Rücker and by Tolman and try to explain first what kind of reasoning can lead to a smaller answer, and then what kind of reasoning can lead to a greater answer. For the special purpose of this discussion I shall begin with the five quantities :—

Length, time, energy, electric potential, temperature.

If the required number of fundamental quantities is less than five, it must be possible to express the dimensions of at least one of the above five quantities in terms of the others. It has been pointed out by Dingle‡ that whatever kind of clock a physicist uses, its calibration depends ultimately on defining equal intervals of time as times taken, either by an undisturbed particle or by light in a uniform medium, to travel equal distances. Closely related to this is the statement that any time interval t can be correlated to a definite length l by the relation

$$\mathbf{c}t=l. \qquad (18.1)$$

By means of this relation any time interval t can, if desired, be measured by quoting the value of the related length l. Therefore, say some authors, time and length have the same dimensions. Let us provisionally accept this dubious reasoning, pursue it to its limit and see where it leads us.

Consider an amount of energy E. If now we divide E by the universal constant $\mathbf{h}$, we obtain $E/\mathbf{h}$, having the same dimensions as frequency or time^{-1}. Similarly, $E/\mathbf{hc}$ has the same dimensions as length^{-1} and $E/\mathbf{k}$ has the same dimensions as a temperature. Finally, $E/\mathbf{e}$ has the same dimensions as an electric potential. Thus the kind of reasoning that leads to the same dimensions for time and length can also lead to the same dimensions for all five quantities—length^{-1}, time^{-1}, energy, electric potential and temperature. In other words, it would be possible to express the dimensions of any physical quantity whatever in terms of a single dimension. This conclusion is, perhaps, not absurd, but, if it is accepted, there remains almost nothing of dimensional analysis, and we might as well give up altogether the concept of dimensions.

---

* Rücker, Proc. Phys. Soc. x. p. 37 (1888).

‡ Tolman, Phys. Rev. ix. p. 237 (1917).

‡ Dingle, 'The Special Theory of Relativity,' pp. 37–39. Methuen (1940).

As mentioned in § 5, dimensional analysis makes use of the necessary conditions that the two sides of any equation must have the same dimensions. If the number of fundamental quantities is artificially reduced below the value five, the power of dimensional analysis is reduced. This is made admirably clear by Tolman *.

It would thus appear that any author is entitled to reduce below five the number of fundamental quantities having independent dimensions, if it pleases him, but any apparent gain in symbolic elegance will be offset by a loss in usefulness of the concept of dimensions.

We turn now to the question whether the number of fundamental quantities with independent dimensions may not be greater than five. I think the answer depends on the type of problem with which one is concerned. Tolman explains how the extension of the field of study from that of space to kinetics requires a new fundamental quantity (time); the further extension to mechanics requires another (mass or energy); the extension of the field to electric phenomena requires another (electric charge); finally, the extension of the field to thermal phenomena requires still another (temperature or entropy).

This argument certainly seems reasonable, but let us return to the starting-point of the argument. It has been taken for granted that lengths in different directions have the same dimensions. This seems the only reasonable view to adopt in general, but in particular problems some alternative view may be at least as reasonable. Suppose a branch of physics were devoted to railway tracks. Length of track and width of track would never enter any problem or formula in a similar way, and there would be no compelling reason for assigning the same dimensions to lengths in the two orthogonal directions.

A more realistic example is afforded by the study of flow of liquid through tubes or flow of electricity along wires. The important characteristics of a tube or wire are their lengths and cross-sections. The fact that a cross-section can, if desired, be considered to have the same dimensions as a length squared is here largely irrelevant. The power of dimensional analysis as applied to such problems is, in fact, increased by treating cross-section and length as independent fundamental quantities.

We thus see that for special problems it may be advantageous to increase the number of fundamental quantities above the usual number.

---

* *Loc. cit.* p. 250.

It can sometimes be useful in dimensional analysis to regard the *number of atoms* as having dimensions different from a pure number *.

I admit that the views expressed in this concluding section are controversial, and I trust that they will not detract from the rest of the article. If they are sound, then the number of fundamental quantities having independent dimensions is, to some extent, a matter of choice. But, if in the same problem or set of problems two authors make a different choice, the one choosing the greater number is likely to be the more competent physicist.

Department of Chemical Technology,
Imperial College, London.
March 1942.

L. *Contact Potentials.*—Part III. *The Voltaic Cell.*

By J. Alan Chalmers, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham †.

[Received March 20, 1942.]

1. *Introduction.*

In Part I. we have laid down general principles governing the subject of contact potentials, and in Part II. these have been applied to the discussion of the Volta effect. Now we turn to the question of the voltaic cell, which is more complex than the Volta effect, since we now have changes of carrier.

In the simple voltaic cell, we have a system consisting of two metals and one electrolyte. But in many types of voltaic cell, for example, the Daniell cell, there are two electrolytes present, involving the consideration of effects at the junction of two electrolytes. It would be possible to consider the potential difference of such a system to be made up of the potential differences of two systems each containing only three phases, but this would involve the discussion of effects at a boundary between a metal and an electrolyte not containing the ion of the metal, as, for example, if we consider the potential difference of the Daniell cell to be the sum of the potential differences of the systems: $Cu/Zn/ZnSO_4/Cu$ and $Cu/ZnSO_4/CuSO_4/Cu$. The question of the $ZnSO_4/Cu$ boundary (the effect of which cancels out on adding the potential differences) would make such systems difficult to discuss, and it is preferable to consider a cell such as the Daniell cell as a system of four phases. It is a simplification to consider first a cell in which each electrolyte contains the positive ion of the metal with which it is in contact, and no other positive ions in appreciable numbers (unless the solution is very dilute, the H^+ ions due to water can be neglected). We shall consider separately the metal-electrolyte junctions and the junction of the two electrolytes and then add these to the results for the metal-metal junction obtained in Part II. As discussed in Part I., the total potential difference of the cell can be obtained by adding either the internal or the external potential differences.

---

* See, for example, Fowler and Guggenheim, 'Statistical Thermodynamics,' p. 384 (1939).

† Communicated by the Author.

In discussing the cell, we must consider it not to be giving a continuous current but, rather, to be in a potentiometer circuit with a balancing potential applied, so that there will be equilibrium conditions. In discussing the equilibrium conditions we have to realise that, in some cases, the carriers of the current across a boundary differ according to the direction in which the current is flowing, and both have to be discussed; this occurs, for example, at a junction of two electrolytes.

2. *The Metal-Electrolyte Boundary.*

At a metal-electrolyte boundary, there is a tendency for metal ions to pass from the metal into the electrolyte, and for metal ions from the solution to be deposited on the metal itself. In addition, we must consider the effect of negative ions reaching the boundary; when the negative ions give up their charges the uncharged atom or acid radicle combines with a metal atom, giving a molecule of the salt which then dissociates into ions; the net result of this process is thus the passage of a metal ion into solution. Thus, the effect of negative ions reaching the boundary brings fresh metal ions into solution, whereas the effect of positive ions reaching the boundary is to deposit metal. Since it is only metal ions which cross the boundary, the condition of equilibrium is that there is no change of thermodynamic potential when a metal ion moves either way across the boundary. If ${}_3D_1$ is the removal potential difference for metal ions in the metal and ${}_1D_3$ in the solution, we have, from the results of Part I., the external potential difference $A_{13}={}_1D_3-{}_3D_1$, M_{13} being zero, as there is equilibrium.

Now, the carrier of charge in the metal is an electron, with removal potential difference D_1, so that, on the metal side of the boundary there is a carrier change potential difference ${}_3C_1=D_1-{}_3D_1$. If D_3 is the "composite" removal potential difference for the carriers in the solution, ${}_1C_3=D_3-{}_1D_3$ and so we have the internal potential difference,

$$B_{13}=D_1-D_3+A_{13}=D_1-{}_3D_1-D_3+{}_1D_3.$$

If P_3^+, P_3^- are the changes of potential in removing, respectively, a positive and a negative ion from the solution, then ${}_1D_3=P_3^+$ since the current across the boundary is carried by positive ions; taking into consideration that the current is carried by ions of both signs in proportion to their velocities, we obtain the "composite" removal potential difference:

$$D_3=\frac{v_3^-P_3^-+v_3^+P_3^+}{v_3^-+v_3^+},$$

2 K 2

where v_3^+, v_3^- are the mobilities of the positive and negative ions, so that

$$ {}_1C_3 = D_3 - {}_1D_3 = \frac{v_3^-(P_3^- - P_3^+)}{v_3^- + v_3^+}. $$

It is important to note that, using the sign convention as given in part I., P_3^- will be positive when P_3^+ is negative.

We shall leave until later the discussion of the calculation of the D's.

3. *The Junction of Two Electrolytes.*

When we consider the passage of charge across the boundary between two electrolytes, we must take into account that the carriers of charge across the boundary are different according to the direction of the current; we must consider unit potential difference to be applied in each direction and then take the average of these. For notation, let P_3^+, P_3^- be the changes of potential in removing, respectively, the positive and negative ions in the electrolyte 3 from this electrolyte and let P_4^+, P_4^- be the changes of potential in removing these ions from electrolyte 4 when they get there. Similarly, let Q_4^+, Q_4^-, Q_3^+, Q_3^- be the changes of potential in removing the ions of the electrolyte 4. Let n_3, n_4 be the numbers of ions in unit volumes of the two electrolytes, and v_3^+, v_3^-, v_4^+, v_4^- the mobilities of the ions.

If a unit potential difference is applied from the direction of phase 3 to phase 4, the number of ions reaching the boundary in unit time will be $n_3v_3^+$ positive ions from phase 3 and $n_4v_4^-$ negative ions from phase 4. Similarly, reversing the direction of the potential difference, the numbers will be $n_3v_3^-$ negative ions from phase 3 and $n_4v_4^+$ positive ions from phase 4; the changes of potential in each case will be given by the corresponding P or Q. Thus, for the motion of $n_3v_3^+ - n_4v_4^- + n_3v_3^- - n_4v_4^+$ ions from phase 3 the total change of potential in phase 3 is given by

$$ n_3v_3^+P_3^+ - n_4v_4^-Q_3^- + n_3v_3^-P_3^- - n_4v_4^+Q_3^+, $$

and ${}_4D_3$ is given by the change of potential per ion, so

$$ {}_4D_3 = \frac{n_3v_3^+P_3^+ - n_4v_4^-Q_3^- + n_3v_3^-P_3^- - n_4v_4^+R_3^+}{n_3v_3^+ - n_4v_4^- + n_3v_3^- - n_4v_4^+}. $$

Similarly, the composite removal potential difference, ${}_3D_4$, for the carriers of charge from phase 4, will be obtained by using suffix 4 in place of suffix 3 and *vice versâ*. When ${}_4D_3$ and ${}_3D_4$ are known, and assuming M_{34} to be zero, since there is electrical equilibrium, we get

$$ A_{34} = {}_3D_4 - {}_4D_3 $$

giving the external potential difference, and

$$ B_{34} = D_3 - D_4 + A_{34} = D_3 - D_4 + {}_3D_4 - {}_4D_3 $$

giving the internal potential difference.

The question of the actual calculation of the potential differences ${}_3D_4$, ${}_4D_3$ will be left until later.

4. *The Complete Cell.*

We can obtain the potential difference of the complete cell by adding together the internal or external potential differences. The internal potential differences, B_{13}, B_{34}, B_{42} and B_{21} are, respectively,

$$D_1-{}_3D_1-D_3+{}_1D_3, \quad D_3-D_4+{}_3D_4-{}_4D_3, \quad -D_2+{}_4D_2+D_4-{}_2D_4$$

and zero, giving a total potential difference of

$$D_1-{}_3D_1+{}_1D_3+{}_3D_4-{}_4D_3-D_2+{}_4D_2-{}_2D_4.$$

The external potential differences, A_{13}, A_{34}, A_{42} and A_{21} are, respectively, ${}_1D_3-{}_3D_1$, ${}_3D_4-{}_4D_3$, $-{}_2D_4+{}_4D_2$ and D_1-D_2, giving the same total potential difference, but divided differently.

We can notice the part D_1-D_2, which is the Volta effect of the two metals, and it turns out that this is the most important part of the whole potential difference, thus giving rise to the similarity in the results for the Volta effect and voltaic cell for the same pair of metals.

Just as in the case of the Volta effect, so in the voltaic cell there have been the two theories, the contact and chemical theories, which can be interpreted in terms of the external and internal potential differences respectively. In the case of the contact theory, the most important part of the total potential difference occurs in the term A_{21}, while in the case of the chemical theory there are important terms B_{13} and B_{42}.

Since the energy supplied to the voltaic cell is mainly chemical energy, it seems natural to look for the main source of the potential difference at the sources of the chemical energy, namely at the metal-electrolyte boundaries, so that the chemical theory has received much support from this point of view. Unfortunately, very many attempts to discuss potential differences in the voltaic cell have considered what are really the external potential differences at the metal-electrolyte boundaries and have then used the arguments of the chemical theory to assume zero potential difference at the metal-metal junction. It is only comparatively recently that it has been realised that there is a " potential difference " at the metal-metal junction, meaning by this what we have called the external potential difference. It would be possible to divide up the total potential difference of the cell into parts according to the internal potential differences and so have no effect due to the metal-metal junction, but this would not agree with the usual ideas of the " electrode potentials," which are more closely connected with the external than with the internal potential differences.

5. *Elimination of Liquid Junction Potentials.*

Owing to the difficulty of getting reproducible effects for liquid junction potentials, and the fact that such potentials cannot be easily divided into parts corresponding to the two electrodes, it has been thought desirable to attempt to eliminate liquid junction potentials. This has been achieved by a method first used by Tower$^{(1)}$ with a " salt bridge " between the two electrolytes. The salt bridge consists of a concentrated solution of a suitable electrolyte, so arranged that its diffusion into the two electrolytes is slow and can be neglected during the period in which the observations are made. The electrolyte used is one which contains two ions of equal mobility, so that in the diffusion process into the other electrolytes ions of the two signs carry charges equally and there is no transfer of charge across either boundary of the salt bridge; ammonium nitrate and potassium chloride are found to be suitable electrolytes; since the salt bridge is concentrated, the ions from the other electrolytes carry so small a portion of the current as to be negligible. It is necessary to enquire whether it is the internal or external potential difference which is eliminated by the salt bridge, and we can do this by considering the results previously obtained for the junction of two electrolytes. Using the suffix "5" to represent the salt bridge, we have to find whether $A_{35}+A_{54}$ or $B_{35}+B_{54}$ is zero.

In the previous results for ${}_4D_3$, etc., we can see that n_5 is much larger than n_3 or n_4 and $v_5^+=v_5^-$; using R_3^+, etc., for the properties of the salt bridge ions in the various solutions, we see that

$$ {}_5D_3=\tfrac{1}{2}\{R_3^-+R_3^+\} $$

and

$$ {}_5D_4=\tfrac{1}{2}\{R_4^-+R_4^+\}. $$

Similarly,

$$ {}_3D_5=\tfrac{1}{2}\{R_5^-+R_5^+\}, $$

$$ {}_4D_5=\tfrac{1}{2}\{R_5^-+R_5^+\}. $$

Thus

$$ A_{35}={}_3D_5-{}_5D_3=\tfrac{1}{2}\{-R_3^- -R_3^+ +R_5^- +R_5^+\} $$

and

$$ A_{45}={}_4D_5-{}_5D_4=\tfrac{1}{2}\{-R_4^- -R_4^+ +R_5^- +R_5^+\}. $$

Since R_3^- and R_4^- are the removal potential differences for the negative ions of the salt bridge electrolyte when in the other two solutions, where the concentrations of such ions are small, these can well be expected to be equal, so we see that $A_{35}=A_{45}$ and thus the two ends of the salt bridge together give a zero external potential difference.

Considering now the internal potential differences B_{35} and B_{45},

$$ B_{35}=D_3-D_5+A_{35} $$

$$ B_{45}=D_4-D_5+A_{45}, $$

where $D_5=\tfrac{1}{2}\{R_5^+ +R_5^-\}={}_3D_5={}_4D_5$, since the current is carried by the ions of the intermediate electrolyte both in the salt bridge and across its boundaries.

Thus

$$ B_{35}-B_{45}=D_3-D_4, $$

which is not, in general, zero, and so the salt bridge does not eliminate the internal potential differences at its boundaries.

With a salt bridge the total potential difference of the cell becomes

$$ D_1-{}_3D_1+{}_1D_3-D_2+{}_4D_2-{}_2D_4. $$

6. *Electrode Potentials.*

On the idea that the potential differences are to be located at the electrodes only, it has been the practice to consider that the total potential difference of a cell can be divided into two " electrode potentials " due to two " half-cells " ; if we can assume a value for the electrode potential for one standard half-cell, then the measurement of the total potential difference of a cell composed of this standard half-cell, a salt bridge and some other half-cell gives the electrode potential of the second half-cell ; and, as can be understood from the general principles of the addition of potential differences (see Part I.), the sum of the electrode potentials for any pair of half-cells gives the potential difference of the complete cell with a salt bridge present. From the practical point of view of predicting the potential differences of cells, without the need of measuring all possible cells, the electrode potentials are very convenient and work out quite satisfactorily ; this is the more so since, as we shall see later, the effects of change of concentration can be included in the discussion, so that all that is needed is to tabulate the electrode potentials of the metals in contact with, say, normal solutions of salts, and to know from theoretical considerations the effect of concentration changes.

But it must be clearly realised that electrode potentials are not either internal or external potential differences and cannot be so, however one may choose the electrode potential of a standard half-cell. If it were possible to measure the external potential difference due to one half-cell and then to use this as the electrode potential for the half-cell, the electrode potential of another half-cell would not be the external potential difference of the junction in this half-cell because, though there is no external potential difference at the salt bridge, there is an external potential difference at the metal-metal junction and so the electrode potential would, in this case, be the sum of the external potential differences at the junction in the half-cell and at the metal-metal junction. If, on the other hand, internal rather than external potential differences are to be discussed, then, though there is no internal potential difference at the metal-metal junction, there is now one at the salt bridge, as we have seen in the last section. Thus the " electrode potential " of electrochemistry cannot be either of the potential differences discussed in the present work, and must not be confused with them.

In actual practice, different values have been assumed for standard electrode potentials. For reasons which will be discussed in a later part, it has been considered likely that there is a potential difference, which can be identified with an external potential difference of 0·56 volts at the junction of a normal calomel electrode, consisting of mercury, mercurous chloride paste and normal potassium chloride, the mercury being positive. Using this value for the electrode potential of the normal calomel electrode, it has been wrongly considered that other electrode potentials would also give the " single potential differences " at the electrodes. In view of doubts whether the value of 0·56 volts really represents the " single potential difference " of the electrode, it was suggested that the electrode potential of the normal calomel electrode or some other electrode should be conventionally taken as zero. The most common practice now is to take as zero the electrode potential of the normal hydrogen electrode, consisting of hydrogen at a pressure of 1 atmosphere adsorbed in palladium or platinum, in contact with a solution containing hydrogen ions at unit "activity," corresponding, for a completely ionised ideal solution, with a normal solution. With this conventional zero, the electrode potential of the normal calomel electrode is 0·2812 volts. It is clear that a change from one conventional zero to another only introduces a numerical change of the same magnitude in all the electrode potentials.

Since, as we have seen, the electrode potentials cannot correspond to external potential differences, it is clearly misleading to use the value of 0·56 volts for the calomel electrode.

7. *Calculation of the D's.*

If we consider the ions in solution to behave like particles of a perfect gas, we can calculate their thermodynamic potentials, and hence the D's.

Considering an electrolyte containing two singly charged ions, the internal energy of a negative ion is $E+\frac{3}{2}kT$, where E is the energy required to remove the negative ion from the solution and k the gas constant per ion. It can easily be shown that the entropy per ion is given by

$$S=\frac{3}{2}k\log T-k\log\frac{n}{v}+S_0,$$

where n is the number of ions in a volume v and S_0 is the constant of integration in the entropy equation.

The thermodynamic potential per ion of the negative ions, which has been denoted by $-P_3^-$ is thus

$$E_-+\frac{3}{2}kT+kT-\frac{3}{2}kT\log T+kT\log\frac{n}{v}-S_0T,$$

which can be written as $N_3^-+kT\log\frac{n}{v}$,

where N_3^- is a constant for a given electrolyte, independent of the concentration. So

$$P_3^-=-N_3^--kT\log\frac{n}{v}.$$

Similarly, since P_3^+ is negative if P_3^- is positive,

$$P_3^+=N_3^++kT\log\frac{n}{v}.$$

Now, as we have seen, ${}_1D_3=P_3^+$ and

$$D_3=\frac{v_3^-P_3^-+v_3^+P_3^+}{v_3^-+v_3^+}.$$

Thus $$A_{13}={}_1D_3-{}_3D_1=N_3^++kT\log\frac{n}{v}-{}_3D_1,$$

so that the variation with concentration of the external potential difference is brought in by the term $kT\log\frac{n}{v}$.

And $$B_{13}=D_1-D_3+A_{13}={}_3C_1-{}_1C_3$$
$$={}_3C_1+\frac{v_-}{v_++v_-}\left(N_3^-+N_3^++2kT\log\frac{n}{v}\right).$$

At the junction of two electrolytes, the calculation of the D's is difficult, for such quantities as P_4^+, Q_3^+, etc., must involve terms with $kT\log\frac{n}{v}$, where $\frac{n}{v}$ is the concentration of the ions of electrolyte 3 in the electrolyte 4; results can only be obtained by making assumptions as to the nature of the boundary. In the particular case where there are similar ions on both sides of the boundary, *i. e.* the concentration cell, it is easy to obtain results, and these will be considered in Part IV., which will deal with the concentration cell.

It is to be noted that the above results for A_{13} and B_{13} deal with single ions and the constant k is the gas constant per ion or molecule. If we now deal with the concentration in gm. mols. per litre, we have to bring in an extra term in N_3^+, and the constant k must be replaced by $\frac{R}{F}$, where R is the gas constant per gm. mol. and F the faraday, the charge carried by the positive ions of 1 gm. mol. The quantity $\frac{n}{v}$ then becomes c, the concentration, or if the ionisation is not complete, the "activity," *i. e.* the equivalent concentration for a solution which is completely ionised and in which the ions behave as a perfect gas.

The changes required if the ions are multiply charged, or if there are more than two ions per molecule are easily inferred, and we shall not deal with them here.

8. *Solution Pressures.*

Nernst (2) discussed the problem of the electrode potentials from the idea that when there is equilibrium at a boundary the pressure of the ions (osmotic pressure) must be balanced by the pressure of metal ions tending to go into solution. He obtained the result of a term $\frac{RT}{F}\log c$ for the variation of electrode potential with concentration; using measured electrode potentials, with normal calomel $=0{\cdot}56$, and assuming that these are the electrode potentials of Nernst's theory, the solution pressures have been calculated, and range from 10^{44} atmospheres for Mg to 10^{-15} atmospheres for Hg and Ag. As can easily be seen, these results will be completely vitiated if the electrode potential as measured is not the same as considered in Nernst's theory; this must be what we have termed the external potential difference, and as there is an external potential difference at a junction of, say, Mg and Hg, the electrode potential at a Mg-solution junction is not the external potential difference.

The application of Nernst's theory in the case of a mercury-solution junction does not involve a metal-metal junction, but there still appears to be some difficulty, since mercury in contact with an acid or a salt acquires a positive charge, as can be seen from the results by the capillary electrometer, etc. According to the Nernst theory, this must be due to the deposition of positively charged mercury ions from the solution on to the surface of the mercury ; but this presupposes a certain, even if small, concentration of mercury ions already in the solution. This point will be discussed in more detail in a later part in connection with single potential differences.

9. *Surface Layers.*

The potential differences of voltaic cells do not depend on the natures of the metal surface layer, provided these do not affect the chemical processes of the cells ; this has been taken to indicate that there is no real connexion between the potential difference of a voltaic cell and the Volta effect for the same metals, because the latter is much affected by surface layers. But, as can be seen by reference to Part I., the general principles of contact potentials show that surface layers at a metal-electrolyte boundary do not influence the potential difference of the cell unless they affect the chemical processes there. If there is a surface layer at a metal-electrolyte junction this introduces a new phase and there will now be two boundaries ; the sum of the external potential differences for these two boundaries must be equal to the external potential difference for the pure metal and the electrolyte, otherwise the surface layer would affect the whole potential difference of the cell ; but, if the external potential difference at the junction is measured as a "single potential difference," this may be only one of the two external potential differences and so may not give the effect for the pure metal and electrolyte.

10. *Inert Electrodes.*

At an inert electrode, no metal ion passes into the solution, but, instead, an electron passes across the boundary, either combining with a positive ion to form a neutral atom, or being released from a negative ion forming a neutral atom. In the discussion of such cases, we can use the same principles as before with the difference that it is now the electron which carries the current across the boundary, so that there is no change within the metal. We would require ${}_1D_3$ the removal potential difference of the electron in the electrolyte, which must be the difference between the removal potential differences of, for example, the positive ion and the neutral atom. When, as in the above case, there are two processes at work, it is necessary, of course, to use "composite" potential differences.

Very similar to the above cases will be oxidation-reduction effects at an inert electrode, *e. g.* Fe^{++} and Fe^{+++} ions together in solution. Here the electron converts Fe^{+++} to Fe^{++} and ${}_1D_3$ must be the difference between the removal potential differences of these.

11. *The Gibbs-Helmholtz Equation.*

The Gibbs-Helmholtz equation relating the potential difference of a cell with the chemical change occurring is

$$\Delta G=\Delta H+T\left(\frac{\partial\Delta G}{\partial T}\right)_p,$$

where ΔG is the potential difference and ΔH the chemical change. $T\left(\frac{\partial\Delta G}{\partial T}\right)_p$ represents the energy supplied as heat.

At a single junction, as we have seen, the external potential difference A_{13} is just equal to the difference in thermodynamic potential across the junction, so that we can apply the Gibbs-Helmholtz equation to each junction separately, if we divide up the total potential difference according to external potential differences.

The energy supplied as heat is $T\left(\frac{\partial\Delta G}{\partial T}\right)_p$ or $T\Delta S$, and it is clear that there will be a part of this at each junction. When a current passes across a metal-metal junction, heat is absorbed or given out by the Peltier effect and so, by an extension of the term, we can consider the heating effects in a voltaic cell to be the manifestation of Peltier effects, and the Peltier effect at any junction is $T\frac{\partial A_{13}}{\partial T}$. We may notice that the effect of concentration on the potential difference A_{13} occurs entirely in the term $T\frac{\partial A_{13}}{\partial T}$, as can be seen by reference to the calculation of the D's. Further discussion of the Gibbs-Helmholtz equation is reserved for Part IV., dealing with the concentration cell.

12. *Conclusion.*

We have seen that the phenomena of the voltaic cell can be interpreted by the general principles that have been laid down in Part I. and used in Part II. to interpret the Volta effect. The use of external rather than potential differences is in better agreement with general ideas, but this involves the existence of a potential difference at the metal-metal junctions. Electrode potentials, though useful in obtaining potential differences for cells, are not either external or internal potential differences.

References.

(1) Tower, *Zeit. Phys. Chem.* xx. p. 198 (1898).
(2) Nernst, *Zeit. Phys. Chem.* ii. p. 163 (1888).

LI. *Contact Potentials.*—Part IV. *The Concentration Cell.*

By J. Alan Chalmers, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham *.

[Received March 20, 1942.]

1. *Introduction.*

In Part I. general principles regarding contact potentials have been laid down, and these have been applied in Part II. to the Volta effect and in Part III. to the voltaic cell ; now we turn to a special type of cell, the concentration cell, which shows particular points of interest.

The simple concentration cell consists of two samples of the same electrolyte, but with different concentrations, having a liquid junction and joined together by the metal which is the cation, *e. g.*,

$$\mathrm{Ag}\,|\mathrm{AgNO_3}(c_1)\,|\mathrm{AgNO_3}(c_2)\,|\mathrm{Ag}.$$

It is possible to eliminate the effect of the liquid junction by the use of a salt bridge, as described in Part III. There are also the concentration cell without transference, the salt bridge being replaced by some system such as AgCl paste | Ag | AgCl paste, and the amalgam concentration cell in which there are two amalgams with different concentrations of dissolved metal in the mercury. We shall not concern ourselves with these last two types of concentration cell, nor shall we refer in much detail to the cell with the salt bridge, but we shall mainly confine our attention to the ordinary concentration cell. The example given above has the advantage, for theoretical consideration, that it contains a uni-univalent electrolyte, so there are no effects due to double charges.

2. *The Junction of the Two Electrolytes.*

Using the results obtained in Part III., we can get the removal potential differences ${}_4D_3$ and ${}_3D_4$ and hence $A_{34}={}_3D_4-{}_4D_3$.

$${}_4D_3=\frac{n_3v_3^+P_3^+-n_4v_4^-Q_3^-+n_3v_3^-P_3^--n_4v_4^+Q_3^+}{n_3v_3^+-n_4v_4^-+n_3v_3^--n_4v_4^+},$$

where n_3, n_4 are the number of ions per unit volume, v_3^+, v_3^-, v_4^+, v_4^- the mobilities of the ions, and P_3^+, P_3 the removal potential differences for the ions of electrolyte 3 in that electrolyte, Q_3^+, Q_3^- being the same quantities for the ions of electrolyte 4 when in electrolyte 3.

For the concentration cell, the ions in both electrolytes are the same, so that $v_4^+=v_3^+$, $v_4^-=v_3^-$, $Q_3^+=P_3^+$ and $Q_3^-=P_3^-$.

---

* Communicated by the Author.

Then $${}_4D_3=\frac{v_3^+P_3^+ + v_3^-P_3^-}{v_3^+ + v_3^-}$$

and $${}_3D_4=\frac{v_3^+P_4^+ + v_3^-P_4^-}{v_3^+ + v_3^-}.$$

Thus $$A_{34}=\frac{v_3^+(P_4^+ - P_3^+) + v_3^-(P_4^- - P_3^-)}{v_3^+ + v_3^-}.$$

Now, also from Part III., we have

$$\begin{aligned}P_3^+&=N_3^+ + kT\log c_3,\\ P_4^+&=N_3^+ + kT\log c_4,\\ P_3^-&=-N_3^- - kT\log c_3,\\ P_4^-&=-N_3^- - kT\log c_4,\end{aligned}$$

where N_3^+, N_3^- are independent of concentration and c_3, c_4 are the concentrations in the two electrolytes.

This gives

$$A_{34}=\frac{v_3^+ - v_3^-}{v_3^+ + v_3^-}kT\log\frac{c_4}{c_3},$$

$$D_3=\frac{v_3^+(N_3^+kT\log c_3) - v_3^-(N_3^- - kT\log c_3)}{v_3^+ + v_3^-},$$

and $$D_4=\frac{v_3^+(N_3^+ + kT\log c_4) - v_3^-(N_3^- - kT\log c_4)}{v_3^+ + v_3^-}.$$

So $B_{34}=D_3-D_4+A_{34}$ is zero, which also follows from the fact that there is no change of carrier and, as there is also no transfer potential difference, there can be no internal potential difference.

3. *The Potential Difference of the Complete Cell.*

From Part III.,

$$A_{13}=N_3^+ + kT\log c_3 - {}_3D_1,\quad A_{41}=N_3^+ + kT\log c_4 - {}_4D_1.$$

But ${}_3D_1={}_4D_1$, each being the removal potential-difference for the metal ion when inside the metal.

So, for the whole cell,

$$\begin{aligned}L_{1341}&=A_{13}+A_{34}+A_{41}\\ &=\frac{2v_3^-}{v_3^+ + v_3^-}kT\log\frac{c_3}{c_4},\end{aligned}$$

the well-known result for the potential difference of the concentration cell, though obtained by a method of argument differing from the usual.

If we have a concentration cell with a salt bridge, the external potential difference A_{34} is eliminated, and we have the potential difference

$$kT\log c_3/c_4.$$

It would seem that the comparison of the e.m.f.'s of concentration cells with and without a salt bridge would give a means of determining the "transport number" without the need of measuring changes of concentration, as in Hittorf's method.

4. *The Gibbs-Helmholtz Equation.*

Using the Gibbs-Helmholtz equation

$$\Delta G=\Delta H+T\left(\frac{\partial \Delta G}{\partial T}\right)_p,$$

we see that $\Delta G=T\left(\frac{\partial \Delta G}{\partial T}\right)_p$, so that, in the concentration cell, the chemical energy supplied (ΔH) is zero. The whole energy of the concentration cell is thus supplied by heat energy, and hence a concentration cell when supplying a current will become colder. At first sight this might seem to be a contradiction of the second law of thermodynamics, energy being derived from the heat of bodies no warmer than the surroundings; but it can be seen that when a concentration cell is giving a current there is a passage of ions across the boundaries in such a way that the difference of concentration between the two solutions tends to diminish; or, to put it in another way, although the cooling of the cell when running amounts to a decrease of entropy, this is offset by a net increase in the entropies of the two solutions, and, if the cell is working reversibly, there is no resultant change of entropy. Owing to the tendency towards equalization of the concentrations, the cell will not continue to give a current permanently.

It is interesting to notice that the fact that the concentration cell gives a resultant potential difference can be associated with the changes of carrier inside the electrolytes, as can be seen by considering the internal potential differences; there is a zero value for B_{34}, while $B_{13}={}_3C_1-{}_1C_3$ and $B_{14}={}_4C_1-{}_1C_4$. Now ${}_3C_1={}_4C_1$ being the carrier change potential difference inside the metal. Thus the total potential difference is made up of ${}_1C_4-{}_1C_3$, the carrier change potential differences at the two ends of the electrolytes. The same result can be seen by considering a case in which there is no carrier change, as can be imagined by considering a concentration cell, not of silver nitrate, but of a silver salt of a very heavy anion, whose mobility will be very small; then silver ions will carry practically all the current in the electrolytes, so that there are no carrier change potential differences in the electrolytes; the ratio $\frac{v_3^-}{v_3^++v_3^-}$ becomes almost zero and, since the total potential difference is proportional to this ratio, this becomes very small. In the silver nitrate cell and in the cell with the heavy anion the current is carried across both metal-electrolyte junctions by silver ions, but it is only in the case of heavy anion that it is carried across the liquid junction also by silver ions alone; in the case of silver nitrate, a part of the current across the liquid junction is carried by the nitrate ions. With the heavy anion, just as many silver ions leave each solutions as enter it, but with the silver nitrate solution, more silver and nitrate ions enter the solution round the anode than leave it, and the reverse occurs for the solution round the cathode; if the former is the dilute solution, then the effect of the current is to give an increase of entropy; this is a natural process and if the passage of charge is reversible, then there will be a balancing decrease of entropy due to the cooling of the system.

In the Gibbs-Helmholtz equation, $\Delta G=\Delta H+T\Delta S$, and we see that, for the whole cell, ΔH is zero, since there is no total chemical change when a charge passes ; in the metal, one silver ion enters and one leaves the metal, and in the solutions it is only $T\Delta S$ and not ΔH which depends on the concentration, so that the effects on ΔH in the two solutions balance out. If we consider each junction separately, ΔH is not zero for each metal-electrolyte junction though the same for both. For the liquid junction, ΔH is zero, for H per particle is $E+pv$ in each solution.

5. *Peltier Effects.*

We have seen that the energy for the concentration cell arises from the cooling of the cell, and the major part, at any rate, of this cooling must occur at the 3 junctions. Now, when a current crossing a junction of two metals causes a heating or cooling effect, this is referred to as a Peltier effect and we can extend this term to heating or cooling effects at metal-electrolyte junctions or liquid junctions.

At any junction, the external potential difference, A, is given by $\Delta G=\Delta H+T\Delta S$, since A represents the electrical potential difference necessary to make the thermodynamic potentials the same on both sides of the boundary. Dividing the whole potential difference of the cell into parts according to the external potential differences, we can consider the " chemical " part of each potential difference as the ΔH, and the " thermal " part as the $T\Delta S$.

Then $T\Delta S$ at any junction is the potential difference of the Peltier effect.

Referring to the results of Part III., we see that TS per metal ion in the electrolyte is

$$\tfrac{3}{2}kT\log T-kT\log c+TS_0.$$

Inside the metal, we can write it as TS_1, where the value of S_1, the entropy per metal ion, can be obtained by integration of the specific heat.

The Peltier effect at the metal-electrolyte junction is thus

$$\tfrac{3}{2}kT\log T-kT\log c+TS_0-TS_1.$$

A measurement of this quantity, if S_1 were calculated, would appear to give a new method of obtaining S_0, the constant of integration of the entropy equation, and so of obtaining absolute entropies for ions in solution.

At the liquid junction, there is no " chemical " part of the potential difference and so the whole external potential difference,

$$A_{34}=\frac{v_3{}^{+}-v_3{}^{-}}{v_3{}^{+}+v_3{}^{-}}kT\log\frac{c_4}{c_3},$$

is the Peltier effect.

A measurement of this would verify the results, or would give a way of obtaining the " transport number.".

We notice that the Peltier effect at the junction of two electrolytes is proportional to the absolute temperature. It will be remembered that the original thermodynamic theory of thermoelectricity, produced by Thomson (later Lord Kelvin), without the inclusion of the Thomson effect, predicted that the thermoelectric e.m.f., due entirely to the Peltier effect, would be proportional to the temperature difference between the junctions. It was the fact that the thermoelectric e.m.f. is not usually directly proportional to the temperature difference that led to the discovery of the Thomson effect.

The revised thermodynamic theory gave the results

$$\pi = \mathrm{T}\frac{d\Theta}{d\mathrm{T}},$$

$$\frac{d\Theta}{d\mathrm{T}} = \frac{d\pi}{d\mathrm{T}} - (\sigma_1 - \sigma_2),$$

where Θ is the thermoelectric e.m.f., π the Peltier effect, and σ_1, σ_2 the Thomson effects in the two conductors.

It is easily seen that if, as is the case for a junction of two solutions of different concentrations of the same electrolyte, π is proportional to T, then $\sigma_1 - \sigma_2$ must be zero. This can be verified, as will be seen in the next section.

A thermoelectric circuit of two solutions of different concentrations of the same electrolyte thus realises the conditions of Thomson's original theory.

6. *The Thomson Effect.*

The Thomson effect is the heat carried along by a current when it moves along a temperature gradient ; this has been identified with the "specific heat of electricity," but Sommerfeld and Frank (1) have shown that this is not correct, and they have defined five different "Thomson coefficients" ; we shall only consider three of these and we shall define them in slightly different ways from those used by Sommerfeld and Frank.

They define the Thomson energy coefficient, μ_{E}, the "true" Thomson coefficient of the thermoelectric circuit, as the coefficient of the term involving both the electric current and the temperature gradient in the expression for the heating produced in a metal. For our purposes, we define σ_{E} as the amount of heat given out when one electron, or negative ion, moves up a unit temperature gradient.

Sommerfeld and Frank define the Thomson heat current, μ_{W}, as the heat carried along a metal at a uniform temperature when there is unit potential gradient. We define σ_{W} as the heat carried along with it by one electron or negative ion.

They define the Thomson potential gradient, μ_{P}, as the potential gradient produced by unit temperature gradient when no current is flowing. We define σ_{P} as the Volta potential difference produced between two portions of the same conductor at unit temperature difference and hence as the difference in the thermodynamic potentials per carrier between the two portions of the conductor.

Sommerfeld and Frank deduced a relation between these three quantities, and, corresponding to this, we get

$$-\sigma_E = \sigma_P + \left(\frac{\partial}{\partial T}(\sigma_W)\right)_c .$$

Now,
$$\sigma_W = \left(\frac{\partial Q}{\partial n}\right)_{c,\,T} = T\left(\frac{dS}{\partial n}\right)_{c,\,T}$$

and
$$\sigma_P = -\left(\frac{\partial}{\partial T}\left(\frac{\partial G}{\partial n}\right)_{p,\,T}\right)_{c,\,n} .$$

So
$$-\sigma_E = -\left(\frac{\partial}{\partial T}\left(\frac{\partial G}{\partial n}\right)_{p,\,T}\right)_{c,\,n} + \left(\frac{\partial}{\partial T}T\left(\frac{\partial S}{\partial n}\right)_{c,\,T}\right)_c ,$$

which gives :

$$\sigma_E = -\left(\frac{\partial}{\partial T}\left[-\frac{\partial G}{\partial n} + T\frac{\partial S}{\partial n}\right]\right)_c$$

$$= \frac{\partial^2 G}{\partial T\,\partial n} - \frac{\partial S}{\partial n} - T\frac{\partial^2 S}{\partial n\,\partial T} .$$

But
$$S = \frac{\partial G}{\partial T} .$$

So
$$\sigma_E = -T\left(\frac{\partial^2 S}{\partial n\,\partial T}\right)_c .$$

Using the results of Part III. for $\frac{S}{n}$, we see that $\sigma_E = -\frac{3k}{2}$, both for positive and negative ions, which is the same for every concentration of electrolyte, so that $\sigma_1 = \sigma_2$ and the results of the last section are verified. If the ions in solution behave as particles with more than 3 degrees of freedom, then $\sigma = -\frac{k}{\gamma - 1}$, where γ is the ratio of the two specific heats.

Richardson ⁽²⁾ gives results for the thermoelectric quantities of metals, which are derived from the assumption that the behaviour of the electrons in the metals is similar to that of gas particles. It has long been realized that the discrepancy between these results and the measured quantities is due to the fact that electrons in metals behave very differently from gas particles. But, when the current is carried by ions in solution, then the behaviour is just like that of gas particles, and the results quoted by Richardson may be expected to hold. In fact, with a few simplifications, the results given above agree with those given by Richardson, and we may say that a thermoelectric circuit composed of two solutions of different concentrations of the same electrolyte realizes the conditions which were assumed, but do not hold in the thermoelectric circuit of metals.

7. *The Concentration Thermojunction.*

The thermoelectric circuit comprised of two junctions, at different temperatures, of solutions of different concentrations of the same electrolyte can be called the "concentration thermojunction," and, as we have seen, the Thomson effects in the two arms will cancel out, so that the resultant e.m.f. is due entirely to the Peltier effect and amounts to $\frac{v_3^+ - v_3^-}{v_3^+ + v_3^-} k(T' - T) \log \frac{c_4}{c_3}$, where T′ and T are the temperatures of the two junctions.

It is clear that the thermoelectric "law of intermediate temperatures" will hold in this case, as for the ordinary thermojunction. But the "law of intermediate metals" does not hold; this states that the insertion of a third metal, 3, between the metals 1 and 2, does not affect the total potential difference of the system if both new junctions, 31 and 32, are at the same temperature as the old junction, 12. Referring to Part I., we see that this follows at once from the result that a system of three metals has zero potential difference. The same "law of intermediate conductors" will be true when some of the conductors are electrolytes only if the system of three conductors has zero potential difference. Now, if the three conductors concerned consist of a metal and the two solutions of different concentrations, the system of three conductors is the ordinary concentration cell, which has a definite, non-zero, potential difference, and so the insertion of a metal in one junction of the concentration thermojunction will effect the potential difference of the circuit, and in this case the "law of intermediate metals" does not hold. If, instead, we insert between the two electrolytes a third electrolyte of concentration different from either of the others, we can find the effect by considering the potential difference of the system of three electrolytes. We can see that this is zero most simply by considering the potential difference of the system to be built up of the internal potential differences, and, as we have seen in section 2, these are zero at each junction of two solutions of different concentrations of the same electrolyte; the same result can, of course, be obtained by considering the external potential differences, which add up to zero; and it is possible to get the result from general principles if we realize that, since the mobilities of the ions are the same in solutions of different concentrations, the effect of the passage of a current round the circuit is that the same number of positive and negative ions enter each solution as leave it, and so the passage of a current produces no change in the circuit; so no chemical energy is available to make the current flow, and if there were a current due to thermal energy, as in the concentration cell, then in this case, unlike the concentration cell, we could get the current flowing permanently without any change in the solutions, and this would provide a contradiction of the second law of thermodynamics. We thus see that a system of three solutions of different concentrations of the same electrolyte gives a zero potential difference, and hence we get a "law of intermediate concentrations." We see that this law means that, in the ordinary concentration cell, the potential difference does not depend on the sharpness or otherwise of the boundary, and therefore various different ways of calculating the general liquid junction potential differences must lead to the same result in the case of the concentration cell, in spite of the differences in assumptions regarding the boundaries.

8. *Conclusion.*

The application of the results for the Voltaic cell, deduced in Part III., to the special case of the concentration cell yield the familiar result, which is independent of the nature of the liquid boundary. The concentration cell is discussed in terms of the Gibbs-Helmholtz equation, and results are obtained for a concentration thermojunction which shows the existence of a " law of intermediate concentrations " but not a law of intermediate metals.

References.

(1) Sommerfeld and Frank, Rev. Mod. Phys. iii. p. 1 (1932).
(2) Richardson, 'The Electron Theory of Metals,' p. 461 (1916).

LII. *The Transverse Vibrations of a Helical Spring with Pinned Ends and no Axial Load.*

By J. Dick, B.Sc., Ph.D.*

[Received November 16, 1941.]

An approximate calculation for the frequencies of transverse vibration of a uniform helical spring can be obtained by applying the well-known mathematical analysis for the transverse vibrations of a long, thin, uniform rod. In this analysis transverse deflexion due to transverse shear forces, and the rotary inertia effects involved, would be neglected. A more complete analysis shows, however, that two distinct types of vibration can occur, and that the approximation applies to only one of these. The distinction between the two types of vibration is that in one case the transverse deflexion due to elastic bending and the transverse deflexion due to shear have the same sign at any instant, and in the other case they have opposite signs, the deflexion due to shear having the greater numerical value. The approximate analysis gives good approximations in the former case, if the distance between adjacent nodes is large compared with the mean diameter of the spring. Each of the two types of vibration has its own " harmonics," the frequency of each depending on the number of nodes.

In the following analysis it is proposed to consider only springs formed from wire of circular cross-section.

List of Symbols used.

F = longitudinal stiffness of spring.
l = length of spring.
D = mean diameter of spring.

---

* Communicated by the Author.

W=weight of spring.

w=weight per unit length of spring$=\frac{W}{l}$.

M=bending moment at any section.

S=transverse shear force at any section.

y_1=transverse deflexion of axis due to bending.

y_2=transverse deflexion of axis due to shear.

$y=y_1+y_2$.

C=shear modulus for transverse shear distortion.

$\frac{1}{m}$=Poisson's ratio for the material$=-\frac{\text{secondary strain}}{\text{primary strain}}$.

$p/2\pi$=frequency of transverse vibration.

The x-axis is taken as passing through the pinned ends of the spring.

Analysis.

The equation for rotary equilibrium of a small element of length δx about an axis perpendicular to the spring axis is

$$\frac{\partial M}{\partial x}\delta x=S\,\delta x+\frac{w}{g}\delta x\left\{k^2+\frac{1}{12}(\delta x)^2\right\}\frac{\partial^2\dfrac{\partial y_1}{\partial x}}{\partial t^2},$$

where $k^2=\frac{D^2}{8}$.

This gives

$$\frac{\partial M}{\partial x}=S+\frac{w}{g}\frac{D^2}{8}\frac{\partial^3 y_1}{\partial x.\partial t^2}. \quad \text{. . . . . (1)}$$

The expression for elastic resistance to bending is obtained by considering the spring equivalent to a beam, for which the value of "EI" can be obtained in terms of the longitudinal stiffness, the length, the mean diameter, and Poisson's ratio With a spring formed from wire of circular cross-section, and using the relationships of the various quantities given by the well-known formulæ for a close-coiled helical spring, it is easy to show that

$$\text{“EI”}=Fl\frac{D^2}{2}\left(\frac{m+1}{2m+1}\right),$$

and on applying the equation for the bending of a beam, assuming small deflexions, this gives

$$M=Fl\frac{D^2}{2}\left(\frac{m+1}{2m+1}\right)\frac{\partial^2 y_1}{\partial x^2}. \quad \text{. . . . . (2)}$$

The transverse shear modulus can also be expressed in terms of the spring stiffness, the length, and Poisson's ratio. Thus

$$C=2Fl\left(\frac{m+1}{m}\right).$$

Hence
$$S=-C\frac{\partial y_2}{\partial x}$$
$$=-2Fl\left(\frac{m+1}{m}\right)\left\{\frac{\partial y}{\partial x}-\frac{\partial y_1}{\partial x}\right\}. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Again, for transverse equilibrium of a specimen element
$$\left(S+\frac{\partial S}{\partial x}.\delta x\right)-S=-\frac{w}{g}\delta x\frac{\partial^2 y}{\partial t^2},$$
and this gives
$$\frac{\partial S}{\partial x}=-\frac{w}{g}\frac{\partial^2 y}{\partial t^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

On eliminating all variables from equations (1), (2), (3), and (4) except the derivatives of y with respect to x and t, and putting $\frac{W}{l}=w$, we have
$$\frac{\partial^4 y}{\partial x^4}-\frac{W/g}{F}\frac{1}{4l^2}\left(\frac{4m+1}{m+1}\right)\frac{\partial^4 y}{\partial x^2\,\partial t^2}+\frac{W/g}{F}\frac{2}{(lD)^2}\left(\frac{2m+1}{m+1}\right)\frac{\partial^2 y}{\partial t^2}$$
$$+\frac{1}{8}\left(\frac{W/g}{F}\right)^2\frac{1}{l^4}\frac{m(2m+1)}{(m+1)^2}\frac{\partial^4 y}{\partial t^4}=0. \quad . \quad . \quad (5)$$

The solution we require is of the form $y=XT$, where X is a function of x only, and T is a harmonic function of t, of frequency $p/2\pi$. Hence
$$\frac{\partial^4 X}{\partial x^4}+\frac{W/g}{F}\cdot\frac{1}{4l^2}\left(\frac{4m+1}{m+1}\right)p^2\frac{\partial^2 X}{\partial x^2}$$
$$-\left(\frac{W/g}{F}\right)^2\frac{2}{(lD)^2}\left(\frac{2m+1}{m+1}\right)\left\{\frac{F}{W/g}p^2-\frac{1}{16}\left(\frac{D}{l}\right)^2\left(\frac{m}{m+1}\right)p^4\right\}X=0, \quad (6)$$
that is,
$$\frac{\partial^4 X}{\partial x^4}+\alpha\frac{\partial^2 X}{\partial x^2}+\beta X=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

If $p^2<\frac{F}{W/g}16\left(\frac{l}{D}\right)^2\left(\frac{m+1}{m}\right)$, then β is negative, and the solution is
$$X=A_1\sin n_1x+B_1\cos n_1x+C_1\sinh n_2x+D_1\cosh n_2x, \quad . \quad . \quad (8)$$
where
$$n_1^2=\frac{\alpha}{2}+\sqrt{\left(\frac{\alpha}{2}\right)^2-\beta}\ . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$
and
$$n_2^2=-\frac{\alpha}{2}+\sqrt{\left(\frac{\alpha}{2}\right)^2-\beta}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

With pinned ends, the terminal conditions are:

when $x=0$, $\quad X=0$ and $\frac{\partial^2 X}{\partial x^2}=0$, and

when $x=l$, $\quad X=0$ and $\frac{\partial^2 X}{\partial x^2}=0$.

These give $B_1=C_1=D_1=0$ and $n_1l=\pi$, 2π, 3π, etc.
Taking the fundamental mode
$$\frac{\alpha}{2}+\sqrt{\left(\frac{\alpha}{2}\right)^2-\beta}=\frac{\pi^2}{l^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

which gives a quadratic for p^2, and the lower value of the two solutions is the one that is valid in equation (11). This is

$$p^2=\left(\frac{m+1}{m}\right)\frac{\mathrm{F}}{\mathrm{W}/g}\left[\left\{8\left(\frac{l}{\mathrm{D}}\right)^2+\pi^2\left(\frac{4m+1}{2m+1}\right)\right\}\right.$$
$$\left.-\sqrt{\left\{\left[8\left(\frac{l}{\mathrm{D}}\right)^2+\pi^2\left(\frac{4m+1}{2m+1}\right)\right]^2-8\pi^4\frac{m}{2m+1}\right\}}\right]. \quad (12)$$

Thus the lower frequency is obtained as a function of $\left(\frac{l}{\mathrm{D}}\right)$ for the spring, and Poisson's ratio for the material.

If $p^2>\frac{\mathrm{F}}{\mathrm{W}/g}16\left(\frac{l}{\mathrm{D}}\right)^2\left(\frac{m+1}{m}\right)$, then β is positive in equation (7), and

$$\mathrm{X}=\mathrm{A}_2\sin n_1x+\mathrm{B}_2\cos n_1x+\mathrm{C}_2\sin n_3x+\mathrm{D}_2\cos n_3x, \quad . \quad . \quad (13)$$

where
$$n_3{}^2=+\frac{\alpha}{2}-\sqrt{\left(\frac{\alpha}{2}\right)^2-\beta}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

The terminal conditions for pinned ends give $\mathrm{B}_2=\mathrm{D}_2=0$, and if A_2 and C_2 are not zero, then $n_1l=\pi$, 2π, 3π, etc., and $n_3l=\pi$, 2π, 3π, etc.

Thus two distinct types of vibration are possible with each mode.

Considering $n_1l=\pi$, the solution for p^2 is that given by equation (12), and it is now clear that the lower value of p^2 given by the quadratic is valid whether it is greater or less than

$$\frac{\mathrm{F}}{\mathrm{W}/g}16\left(\frac{l}{\mathrm{D}}\right)^2\left(\frac{m+1}{m}\right).$$

When $n_3l=\pi$, this gives

$$\frac{\alpha}{2}-\sqrt{\left(\frac{\alpha}{2}\right)^2-\beta}=\frac{\pi^2}{l^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

and
$$p^2=\left(\frac{m+1}{m}\right)\frac{\mathrm{F}}{\mathrm{W}/g}\left[\left\{8\left(\frac{l}{\mathrm{D}}\right)^2+\pi^2\left(\frac{4m+1}{2m+1}\right)\right\}\right.$$
$$\left.+\sqrt{\left\{\left[8\left(\frac{l}{\mathrm{D}}\right)^2+\pi^2\left(\frac{4m+1}{2m+1}\right)\right]^2-8\pi^4\frac{m}{2m+1}\right\}}\right]$$
$$=\left(\frac{m+1}{m}\right)\frac{\mathrm{F}}{\mathrm{W}/g}\left[\left\{8\left(\frac{l}{\mathrm{D}}\right)^2+\pi^2\left(\frac{4m+1}{2m+1}\right)\right\}\right.$$
$$\left.+\sqrt{\left\{64\left(\frac{l}{\mathrm{D}}\right)^2+16\left(\frac{l}{\mathrm{D}}\right)^2\pi^2\left(\frac{4m+1}{2m+1}\right)+\frac{\pi^4}{(2m+1)^2}\right\}}\right], \quad (16)$$

and since this is greater than $\frac{\mathrm{F}}{\mathrm{W}/g}16\left(\frac{l}{\mathrm{D}}\right)^2\left(\frac{m+1}{m}\right)$ whatever the value of $\frac{l}{\mathrm{D}}$, value of p^2 given by this equation is valid.

The distinction between the two types of vibration is made clear if we consider the relative signs of y_1 and y. From equations (3) and (4)

$$\frac{\partial^2y_2}{\partial x^2}=-\frac{1}{2\mathrm{F}l}\left(\frac{m}{m+1}\right)\frac{\partial\mathrm{S}}{\partial x}$$
$$=\frac{\mathrm{W}/g}{\mathrm{F}}\cdot\frac{1}{2l^2}\left(\frac{m}{m+1}\right)\frac{\partial^2y}{\partial t^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

Hence
$$\frac{\partial^2 y_1}{\partial x^2}=\frac{\partial^2 y}{\partial x^2}-\frac{\partial^2 y_2}{\partial x^2}$$
$$=\frac{\partial^2 y}{\partial x^2}-\frac{W/g}{F}\frac{1}{2l^2}\left(\frac{m}{m+1}\right)\frac{\partial^2 y}{\partial t^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (18)$$

For vibrations of the type giving the lower frequency, equation (12),

$$y=\mathrm{AT}\sin n_1 x,$$

where T is a harmonic function of t, of frequency $p/2\pi$. Hence, from equation (18),

$$\frac{\partial^2 y_1}{\partial x^2}=\left\{-n_1{}^2+\frac{W/g}{F}\frac{1}{2l^2}\left(\frac{m}{m+1}\right)p^2\right\}y. \quad . \quad . \quad . \quad . \quad . \qquad (19)$$

If $\frac{n_1{}^2}{p^2}\frac{F}{W/g}2l^2\left(\frac{m+1}{m}\right)$ is greater than unity, the sign of the coefficient of y is negative, and if the quantity is less than unity the sign of the coefficient of y is positive. On substituting for n_1 from equation (9)

$$\frac{n_1{}^2}{p^2}\frac{F}{W/g}2l^2\left(\frac{m+1}{m}\right)$$
$$=\frac{1}{4m}\left\{(4m+1)+\sqrt{\left[(4m+1)^2\right.}\right.$$
$$\left.\left.+\frac{128}{p^2}\frac{F}{W/g}\left(\frac{l}{D}\right)^2(m+1)(2m+1)-8m(2m+1)\right]\right\}$$
$$=\frac{1}{4m}\left\{(4m+1)+\sqrt{\left[1+\frac{128}{p^2}\frac{F}{W/g}\left(\frac{l}{D}\right)^2(m+1)(2m+1)\right]}\right\}. \qquad (20)$$

This is greater than unity, hence

$$\frac{\partial^2 y_1}{\partial x^2}=-a^2 y,$$

which gives
$$y_1=\frac{a^2}{n_1{}^2}y. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (21)$$

Hence y_1 has the same sign as y, and since $\frac{a^2}{n_1{}^2}$ is less than unity, it is obvious that y_2 has also the same sign. It is thus clear that in vibration of the type giving the lower frequency, the transverse shear and bending deflexions have the same signs.

For vibration at the higher frequency, the equation corresponding to (19) is

$$\frac{\partial^2 y_1}{\partial x^2}=\left\{-n_3{}^2+\frac{W/g}{F}\frac{1}{2l^2}\left(\frac{m}{m+1}\right)p^2\right\} \quad . \quad . \quad . \quad . \quad . \qquad (22)$$

and

$$\frac{n_3{}^2}{p^2}\frac{F}{W/g}2l^2\left(\frac{m+1}{m}\right)=\frac{1}{4m}\left\{(4m+1)-\sqrt{\left[1+\frac{128}{p^2}\frac{F}{W/g}\left(\frac{l}{D}\right)^2(m+1)(2m+1)\right]}\right\}, \quad . \quad . \quad . \qquad (23)$$

which is always less than unity; hence

$$\frac{\partial^2 y_1}{\partial x^2}=+b^2 y,$$

which gives
$$y_1=-\frac{b^2}{n_3{}^2}y. \qquad (24)$$

In this case y_1 has the opposite sign to that of y, and hence, also, of y_2, which is greater numerically than y_1.

Summary.

It has been shown that there are two distinct types of vibration. In the first type, the transverse deflexion due to elastic bending and the transverse deflexion due to transverse shear forces have the same sign at any instant, and the frequency f_1 is given by

$$(2\pi f_1)^2=\left(\frac{m+1}{m}\right)\frac{\mathrm{F}}{\mathrm{W}/g}\left\{8\left(\frac{l}{\mathrm{D}}\right)^2+\pi^2\left(\frac{4m+1}{2m+1}\right)\right.$$
$$\left.-\sqrt{\left[\left\{8\left(\frac{l}{\mathrm{D}}\right)^2+\pi^2\left(\frac{4m+1}{2m+1}\right)\right\}^2-8\pi^4\left(\frac{m}{2m+1}\right)\right]}\right\}. \qquad (25)$$

In the second type of vibration, the transverse deflexion due to elastic bending, and the transverse deflexion due to transverse shear forces are always of opposite sign, the shear deflexion having the greater numerical value. The frequency f_2 for this type of vibration is given by

$$(2\pi f_2)^2=\left(\frac{m+1}{m}\right)\frac{\mathrm{F}}{\mathrm{W}/g}\left\{8\left(\frac{l}{\mathrm{D}}\right)^2+\pi^2\left(\frac{4m+1}{2m+1}\right)\right.$$
$$\left.+\sqrt{\left[\left\{8\left(\frac{l}{\mathrm{D}}\right)^2+\pi^2\left(\frac{4m+1}{2m+1}\right)\right\}^2-8\pi^4\left(\frac{m}{2m+1}\right)\right]}\right\}. \qquad (26)$$

The approximate analysis, in which shear deflexion and rotary inertia are neglected, gives the frequency f_0 from

$$(2\pi f_0)^2=\frac{\pi^4}{2}\left(\frac{m+1}{2m+1}\right)\left(\frac{\mathrm{D}}{l}\right)^2\frac{\mathrm{F}}{\mathrm{W}/g}. \qquad (27)$$

For any value of $\frac{l}{\mathrm{D}}$ greater than 5, and any value of $\frac{1}{m}$ less than $\frac{1}{2}$, a close approximation to f_1 is given by

$$f_1=\frac{f_0}{1+2\cdot 33\left(\frac{\mathrm{D}}{l}\right)^2}. \qquad (28)$$

In calculating f_2, however, f_0 is of little value as an approximation. For large values of $\frac{l}{\mathrm{D}}$ an approximation is

$$f_2=\frac{2}{\pi}\frac{l}{\mathrm{D}}\sqrt{\frac{\mathrm{F}}{\mathrm{W}/g}\,\frac{m+1}{m}}, \qquad (29)$$

and if this is expressed in terms of the longitudinal modulus ($\mathrm{K}=\mathrm{F}l$) and the weight per unit length $\left(w=\frac{\mathrm{W}}{l}\right)$, then

$$f_2=\frac{2}{\pi}\frac{1}{\mathrm{D}}\sqrt{\frac{\mathrm{K}}{w/g}\,\frac{m+1}{m}} \text{ approximately}. \qquad (30)$$

Thus for large values of $\left(\frac{l}{D}\right)$ the frequency f_2 is independent of the length, and is a function of the diameter, the modulus, the mass per unit length, and Poisson's ratio.

The frequencies of the "harmonics" for the two types of vibration can be obtained by writing 2π, 3π, 4π, etc., in place of π in the right-hand side of equations (25) and (26). These are not simple multiples of the fundamentals as in the case of the thin uniform rod.

The foregoing analysis applies only to springs having pinned ends and no axial load. The effects of other end constraints, and of axial tension or compression, do not come within the scope of the present article.

LIII. *The Vibratory Electron in Electrostatics.*

By E. Taylor Jones, D.Sc., Professor of Natural Philosophy in the University of Glasgow *.

[Received March 26, 1942.]

In recent papers † the suggestion has been put forward that the electrical properties of stationary electrons and the magnetic properties which they possess when in translatory motion arise from the frequency and energy changes which they experience when interacting as coupled vibrators. The mutual coupling of electrons is regarded as being effected by the electron waves associated with them, so that all electric and magnetic fields appear as modifications of the wave fields which the electrons would have around them when far separated from one another.

Among the systems by which the adequacy of the theory for the explanation of different phenomena has been tested are a considerable variety of groups of stationary electrons (I., II.) and certain magnetic systems consisting of revolving electrons (IV.). In all these cases the theory has been found to lead to results consistent with the accepted laws of electrostatics and magnetism, provided the distance between any two of the electrons is not less than about 10^{-11} cm. At shorter distances the theory indicates a departure from the Coulomb law (I.), and points to a finite and constant value of the mutual energy of two electrons if their distance apart is less than about $2{\cdot}8 \,.\, 10^{-13}$ cm. It has been shown that a positive and a negative electron within this range of each other form a doublet which has the essential properties of a photon (III.), including zero rest-mass, finite energy when travelling at the speed of light, and a transverse electrical axis, and that this type of photon lends itself readily to a satisfactory explanation of the laws of reflexion, refraction (VI.), and dispersion of light in homogeneous transparent media (V.).

---

* Communicated by the Author.

† The papers referred to are as follows, all being in the 'Philosophical Magazine':—I. xxi. p. 337, Suppl. (Feb. 1936); II. xxii. p. 921, Suppl. (Nov. 1936); III. xxiv. p. 458 (Sept. 1937); IV. xxv. p. 682 (April, 1938); V. xxvii. p. 565 (May, 1939); VI. xxxi. p. 394 (May, 1941). The first two papers are on electrostatics, the fourth on magnetism, the third, fifth and sixth on optics.

The earlier papers of the series referred to were concerned with electrostatic problems involving groups of electrons forming regular or symmetrical configurations. In such systems the number of frequencies common to all the members of a group agrees with the number of energy values associated with its different electrical states. For example, if the group consists of three electrons forming an isosceles triangle, the two at the base being opposite in sign, the equations leave the common frequency unaltered if the electrons at the base are interchanged, as they should, since this interchange obviously involves no change in the electrical energy of the system The system therefore has three different energy values and three common frequencies. But if the system is made asymmetrical, *e.g.*, by displacing the electron at the vertex in a direction parallel to the base, it then has four different energy values but still only three common frequencies.

Since the electrostatic energy of a system is, according to the theory, proportional to the change of frequency which the system experiences when its component electrons are brought to their positions in the configuration from the state of infinite separation, it is clear that in asymmetrical systems the theoretical number of energy values represented by the number of common frequencies may be less than the number of values indicated by the Coulomb law. It is the purpose of the present paper to enquire whether and how an asymmetrical system of this kind is able to adapt itself so as to remove this apparent discrepancy, or whether the theory necessarily leads to a departure from the Coulomb law for electrons which are not at extremely short distances from one another.

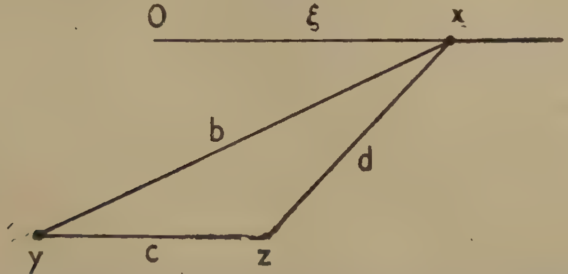
The system chosen for the purpose is that of three electrons forming an equilateral triangle, the one at the vertex being displaced along a line parallel to the base (fig. 1). The equations for the vibratory movements of three electrons in any configuration are

$$\left.\begin{aligned} \ddot{x}+ax+by+dz&=0,\\ \ddot{y}+ay+bx+cz&=0,\\ \ddot{z}+az+dx+cy&=0.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

In these equations x, y, z are the vibratory displacements in the three electrons, these symbols being used in fig. 1 also to " label " the electrons. x is the electron at the vertex, y and z are those at the ends of the base. The coefficient a represents $4\pi^2\nu_0{}^2$, ν_0 being the frequency of an electron at rest and far removed from other electrons. The coefficients b, c, d

represent the coupling of x and y, of y and z, and of z and x respectively. As in the former papers b, c, d are inversely proportional to the distances between the particles the coupling of which they represent.

Fig. 1.



Asymmetrical system of three electrons x, y, z; x is displaced along the horizontal line from O, the vertex of the equilateral triangle Oyz, to infinity.

Assuming $x=Ae^{int}$, $y=Be^{int}$, $z=Ce^{int}$, substituting in equations (1) and eliminating the amplitudes A, B, C, we find for the frequencies the equation

$$N^3-(b^2+c^2+d^2)N-2bcd=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

N being written for n^2-a.

The three values of N given by (2) have been evaluated for various distances of the electron x from the vertex O of the equilateral triangle (fig. 1). This distance being ξ, the distance between the electrons y and z at the base being $l(=k/c)$, and the three roots of equation (2) being N_1, N_2, N_3, the values of ξ/l, b/c, N_1/c, N_2/c, N_3/c are given in the table.

Table.

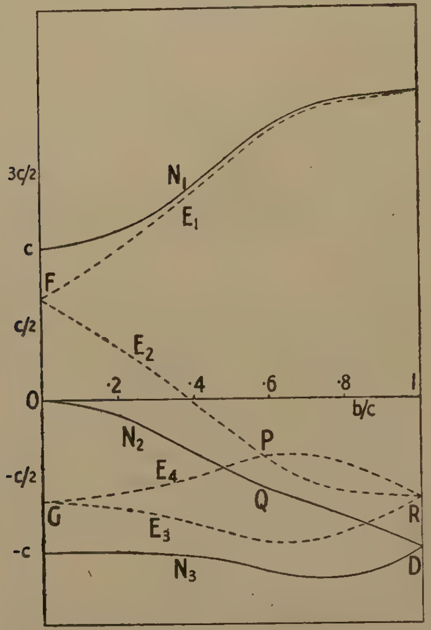
ξ/l.	b/c.	N_1/c.	N_2/c.	N_3/c.
0	1	2	−1	−1
Small.	$1-\xi/2l$	2	$-\left(1-\frac{\xi}{l\sqrt{3}}\right)$	$-\left(1+\frac{\xi}{l\sqrt{3}}\right)$
0·5	0·756	1·949	−0·7409	−1·2081
0·76	0·6541	1·8524	−0·6477	−1·2047
1	0·577	1·732	−0·577	−1·155
1·2	0·5242	1·6312	−0·5193	−1·1119
1·5	0·4588	1·4984	−0·4358	−1·0626
2	0·378	1·3426	−0·3171	−1·0255
3	0·2773	1·1824	−0·1774	−1·005
10	0·0949	1·0196	−0·0194	−1·0002
∞	0	1	0	−1

In all the positions of the electron x the three roots of equation (2) are real, and if c is considerably smaller than a, *i.e.*, if y and z are not nearly at the critical distance r_0, or k/a, from each other, there are also three positive values of n^2, so that the system has three real frequencies.

The values of N_1, N_2, N_3 are also shown, for various values of the coupling ratio b/c, by the full-line curves of fig. 2.

In order to determine what electrical states of the system correspond to the three modes of vitration N_1, N_2, N_3, it is necessary to have the equations for the amplitude ratios, *e.g.*, A/C and B/C. These equations are found from equations (1) to be

Fig. 2.



Frequency and energy curves for asymmetrical system of three electrons.

$$\frac{A}{C}=\frac{Nd+bc}{N^2-b^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$\frac{B}{C}=\frac{Nc+bd}{N^2-b^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

If any value of N_1 (see table) is substituted in these equations, with the corresponding values of b and d, it is found that A, B, and C are (except in the extreme case in which ξ is infinite and b, d, A all zero) all of the same sign, so that the first mode of vibration is that in which the three electrons are vibrating in the same phase and are therefore of the same kind.

In the second mode, for values of ξ between 0 and l, *i.e.*, b/c between 1 and $1/\sqrt{3}$, we find on inserting values of N_2 in equations (3) and (4) that B differs in sign from A and C, so that the electron y (fig. 1) is vibrating in opposite phase to x and z, and therefore differs in kind from them. Over the range of ξ between l and infinity, however, *i.e.*, b/c between $1/\sqrt{3}$ and 0, A is different in sign from B and C, and the electron x is opposite in kind to y and z. In the configuration in which $\xi=l$, represented by the point Q in fig. 2, the triangle xyz (fig. 1) is isosceles, z being equidistant from x and y, and, these being opposite in sign, z has no mutual energy with them. It appears that the second mode of vibration N_2 adapts itself, in different parts of the range of variation of b/c, to the requirements of two electrical states of the system. Over the whole range the electrons x and y are opposite in sign, but in the first part (b/c between 0 and $1/\sqrt{3}$) z is of the same sign as y, in the second part (b/c between $1/\sqrt{3}$ and 1) z has the same sign as x. At the particular value $1/\sqrt{3}$ of b/c the frequency of the system is independent of the sign of z, being given by the equation $n_2{}^2=a-c/\sqrt{3}$.

In the third mode, N_3, equations (3) and (4) show that over the whole range C is opposite in sign to A and B, so that this mode represents an electrical state in which the electron z differs in sign from both x and y.

In any one of the three modes the mutual vibrational energy, calculated on the assumptions that all three electrons are vibrating with the common frequency, and that in each mode N/a is a small fraction, is

$$\begin{aligned}3h(\nu-\nu_0)&=\frac{3h}{2\pi}(\sqrt{a+N}-\sqrt{a})\\&=\frac{3hN}{4\pi\sqrt{a}}\\&=\frac{3Ne^2}{2lc}, \qquad (5)\end{aligned}$$

with e^2 written for $hk/2\pi\sqrt{a}$ as in former papers (*e. g.* I., p. 339), and lc for the coupling constant k. If the mutual vibrational energy is equal to the electrostatic energy E of the system calculated from the Coulomb law, the value of N in each mode of vibration should be equal to the quantity $2lcE/3e^2$ in one of the electrical arrangements of the three particles. Values of this quantity for the four possible arrangements are shown in the broken-line curves in fig. 2. The curve marked E_1 shows values of $2lcE/3e^2$ when all three electrons are of the same kind, E_2 the values when x differs in kind from y and z, E_3 when z is opposite in sign to x and y, and E_4 when y is of different sign from x and z.

The differences between the N and the E curves (fig. 2) can be accounted for in the manner explained in former papers (see I. and II.), viz. :—by the occurrence of states in which the vibrations of the system are fluctuating between modes in one of which all three particles may take part, the other being a partial oscillation shared by only two of them. The occurrence of these fluctuations is indicated in some cases by inequalities of the numerical magnitudes of the amplitudes A, B, C.

The curves N_1 and E_1, for example (fig. 2), agree at one end, where $b/c=1$ and $A/C=B/C=1$ by equations (3) and (4). In this state the particle x is at O (fig. 1), and all three particles are vibrating in the same phase and without fluctuations. But at the other end of curve N_1, where $b/c=0$, or $\xi=\infty$, the ordinate of curve E_1 is two-thirds of that of N_1. In this state the electron x is infinitely far from y and z, b and d are both zero, and N_1 is equal to c (Table I.), so that by (3) and (4) $A/C=0$, and $B/C=1$. The vibrational mode N_1 now affects only two of the three particles, viz., y and z, and the ordinate of the N_1 curve (which is proportional to the mutual energy only if all three particles share equally in the vibration) should be diminished by one-third. This would bring curve N_1 into coincidence with E_1 at the point F (fig. 2).

At intermediate values of b/c, in which this ratio increases from 0 to 1, the amplitude ratio A/C also increases from zero to unity. As explained in previous papers, this is to be interpreted as indicating that the two-particle vibration of frequency $\sqrt{a+c}/2\pi$ does not give place to the three-particle vibration of frequency $\sqrt{a+N_1}/2\pi$ suddenly in one particular configuration, but that the two-particle oscillation persists intermittently, the system fluctuating between the two states, and the duration of the states of partial oscillation gradually diminishing as b/c increases. The other amplitude ratio B/C also is not constant over the whole range. It is unity at both ends, falls to the value $\sqrt{3}/2$ when b/c is $1/\sqrt{3}$ (agreeing with A/C in this configuration, in which the particles x and y are equidistant from z), and falls slightly further to a minimum when b/c is about 0·65. The fact that B is less than C suggests that the partial oscillation is not always confined to the two particles y and z, but that in this part of the range, where x is closer to z than either particle is to y, the partial vibration affects x and z rather than (or alternately with) y and z. The frequency of the xz oscillation would be $\sqrt{a+d}/2\pi$, and, d being less than N_1, this oscillation would also have the effect of lowering the vibrational energy of the system below the value represented by curve N_1 (fig. 2). As suggested in the former papers, the electrostatic energy of the system, represented in the present instance by curve E_1 (fig. 2), is the mean value of the mutual vibrational energy of the system during the fluctuations. It does not of course include the vibrational energy which one of the electrons has in the intermittent periods during the fluctuations when it is vibrating in its undisturbed frequency ν_0. The upper limit of the fluctuations in the present case is represented by curve N_1 (fig. 2), the lower limit by the horizontal line through the point F raised slightly near the right-hand end by the xz partial oscillation, since d is in this part of the range greater than c.

Similar considerations apply to the other curves. The curve E_3 (fig. 2) runs a similar course to that of N_3, but with smaller (negative) ordinates. At the end G, where $b/c=0$, the ordinate of E_3 is two-thirds of that of N_3. In this configuration x is infinitely far from y and z and the mode N_3 (here equal to $-c$) is shared by only y and z, the amplitude of x being zero. The ordinate of curve N_3 at this end should therefore be reduced by one-third for proper comparison with E_3. At the other end R, where $b/c=1$, the triangle formed by the particles is equilateral, and the oscillation N_3 (again equal to $-c$) affects only the two particles at the base, *i.e.*, y and z, which are of opposite kinds. The ordinate at R is therefore, as it should be, two-thirds of that at D. The two-particle vibrations at the ends of curve N_3 evidently affect the whole course of this curve, keeping it below the curve E_3. The limits of fluctuation in this case are the horizontal line GR and the curve N_3. The N_3 vibration is most strongly in evidence near the middle of the curve, the E_3 and N_3 curves being closer together here than they are at the ends.

The part FP of curve E_2 (b/c between 0 and $1/\sqrt{3}$) evidently corresponds to the portion OQ of curve N_2, both representing systems in which the particle x differs in sign from y and z, but the N_2 curve starts from zero and lies below E_2 all along this range. At the point O the particle x is infinitely far from y and z, these two vibrating in their own partial mode of frequency $\sqrt{a+c}/2\pi$, which is not one of the common frequencies of this system. The mutual energy of this partial oscillation is represented by $2\,c/3$, and this addition would raise the N_2 curve at this end into coincidence with curve E_2 at the point F. At the other end Q of this part of the N_2 curve the mutual energy is again confined to two of the particles, x and y in this case, the frequency being now $\sqrt{a-c/\sqrt{3}}/2\pi$, and the ordinate of the point Q should be diminished by one-third in order to represent the energy correctly. This would bring Q into coincidence with P. The two-particle oscillations at the ends of the portion OQ of curve N_2 evidently influence the mutual energy at intermediate points, at which the mode N_2 (three particles) also appears. The electrostatic energy is to be regarded as the mean value of the mutual vibrational energy of the system during the fluctuations among these three possible modes of vibration. The limits of fluctuation for the portion FP of curve E_2 are curve N_2 and the horizontal line through F near the end F, and N_2 and the horizontal line through P near the end P.

As to the remainder of curve E_2, *i. e.*, the lower path between the points P and R, this portion has no part of curve N_2 corresponding to it. It must therefore represent the mean mutual vibrational energy during fluctuations between two partial modes, there being no three-particle vibration within this range. The limits of the fluctuations are evidently the horizontal lines through R and P, the ordinates of which are $-2\,c/3$ and $-2\,c/3\sqrt{3}$. This is the system in which the particle x differs in sign from y and z, and b/c is between $1/\sqrt{3}$ and 1, *i.e.*, ξ/l is between unity and zero.

The course of curve N_2 (fig. 2) undergoes a marked change at Q, and the remainder of it, QD, evidently corresponds to the upper path between P and R (in curve E_4), both representing systems in which the electron y differs in kind from x and z. Both lines are curved the same way, but the curvature of PR (upper path) is greater than that of QD. The differences can be accounted for by the influence of the two-particle vibrations at P and R and the fluctuations between them and the three-particle vibration in the mode N_2. It is also probable that, as suggested in connection with curve N_1, a partial oscillation of x and z, having frequency $\sqrt{a+d}/2\pi$, appears, these being the nearest two of the three particles in this part of the range.

There remains the portion GP of the curve E_4 which, like the lower path PR does not correspond except at its ends with any of the N curves. It represents the electrostatic energy of systems in which the electron y differs in kind from x and z. At G the energy is the same as that in the mode N_3 (two particles), and at P it is equal to that of two particles in the mode N_2. The frequencies in the end states are $\sqrt{a-c}/2\pi$ and $\sqrt{a-c/\sqrt{3}}/2\pi$. Over its course the curve GP appears to represent the mean mutual vibrational energy during fluctuations between these two-particle oscillations, the limiting energy values being the same as those of the lower path between P and R.

The conclusion to be drawn from the vibrational theory of three-electron systems is that the mutual energy of such systems cannot generally be calculated accurately on the assumption that the electrons act upon one another continuously in accordance with the Coulomb law but that in many cases they experience fluctuations between modes of vibration of different frequencies, some of these modes affecting only two of the particles. The existence of such fluctuations is usually indicated in the theory by inequality in the numerical values of the amplitudes. According to the quantum rule two electrons vibrating with the same frequency must have equal energies, and when the theory shows that their amplitudes are different in magnitude this is to be interpreted as indicating that the particles vibrate alternately in two modes, the time spent in one or the other depending upon the ratio of the amplitudes. This ratio approaches one of the values ± 1, as the system comes near to the configuration of regular form in which the mutual energy is given exactly by the expression $\Sigma e^2/r$, e^2 having the value $hk/2\pi\sqrt{a}$, and in which there are no fluctuations. The electrostatic energy of the system in any configuration, and the force acting on one of the electrons, may be regarded as the mean values of the mutual vibrational energy during the fluctuations and of its gradient in the direction of the force. In this way the results of the vibrational theory can be brought into line with those arrived at on the usual assumption that the force between any two electrons of a group is the same as it would be if all the other electrons were absent.

An examination of the table or fig. 2 shows that in certain circumstances one of the frequencies of the three-electron system is imaginary. If, for example, the particles y and z are at the critical distance r_0, or k/a, from each other, their coupling c is equal to a, and the third root N_3 of equation (2) is less than $-a$, except at the ends of the range where $\xi=0$ or ∞ and $N_3=-a$. At all intermediate values of ξ the frequency $\sqrt{a+N_3}/2\pi$ is imaginary, which means that the pair of electrons y and z cannot become coupled with x in these conditions. The system corresponding to the mode N_3 is that in which the electron z differs in sign from x and y. In the present case, c being equal to the limiting coupling a, y and z form a permanent doublet (see III., p. 459) of zero frequency and energy when free from the influence of other electrons and when stationary or moving at a speed less than that of light. In all the configurations corresponding to mode N_3, except at the ends of the range, the electron x is nearer z than y, and, being opposite in sign to z, it would have the effect of lowering the energy of the doublet and causing it to become negative if such coupling were possible. What happens, according to the theory, is that in these circumstances the frequency $n_3/2\pi$ of the whole system is imaginary, so that coupling of the doublet with the electron is impossible. The conclusion, therefore, is that a stationary doublet cannot become coupled with an electron, or in fact a group of electrons, which would have the effect of diminishing the energy of the doublet.

A similar conclusion was arrived at in a former paper for a doublet in motion. When a doublet is moving with the speed of light it has all the properties of a photon, and in the theory of the dispersion of such photons in homogeneous transparent media it was shown (V. p. 569) that, if the velocity of the photon is diminished by a medium through which it is passing, this can happen only if the effect of the medium is to increase the electrostatic energy of the photon. That is, a photon travelling at a speed less than that of light *in vacuo* cannot become coupled with a fixed group of electrons which by its electrostatic action would diminish the energy of the photon. The electrostatic system considered in the present paper, in the mode N_3 with c equal to a, may be regarded as a particular case of this more general result, the velocity of the doublet being here zero.

Natural Philosophy Department,
The University, Glasgow.

LIV. *Heat Conduction in a Wedge, or an Infinite Cylinder whose Cross-section is a Circle or a Sector of a Circle.*

By J. C. Jaeger.*

[Received June 13, 1941.]

1. *Introduction.*

Problems on diffusion in infinite cylinders whose cross-sections are circles or sectors of a circle arise in connexion with heat conduction in, and drying of, whole or split logs of wood. In experimental work on these questions the need has been felt for theoretical results, and the object of this paper is to give the theory of the heating of such solids, with zero initial temperature and constant external temperature, and a collection of numerical results for the most interesting cases. The same problem for the wedge is also considered, both because of its intrinsic interest, and because it is possible to give fairly complete results, since the solution involves only two parameters.

---

* Communicated by the Author.

In § 2 the wedge problem is solved by the use of the Laplace transformation, and in § 3 numerical results are given for wedges of angles 2π, $\frac{1}{2}\pi$, and $\frac{3}{2}\pi$, *i. e.*, a semi-infinite crack and internal and external corners. In § 4 the corresponding problem for the sector of a circle is discussed, and in § 5 numerical results are given for sectors of angles 2π, π and $\frac{1}{2}\pi$. In § 6 the point of the sector of given angle which heats most slowly is determined, and heating curves for these points are given. In §§ 2–6 the solid has been taken to be homogeneous and isotropic; the case of anisotropy arising in wood, namely, different radial and tangential conductivities, is discussed in § 7 and shown to be reducible to the cases previously considered. In § 8 a short table of values for the circular cylinder is given.

2. *Heat conduction in the wedge* $0<\theta<\theta_0$, $r>0$. *The surface* $\theta=0$ *kept at zero and* $\theta=\theta_0$ *at unity for* $t>0$. *The initial temperature zero.*

We have to solve

$$\frac{\partial^2 v}{\partial r^2}+\frac{1}{r}\frac{\partial v}{\partial r}+\frac{1}{r^2}\frac{\partial^2 v}{\partial \theta^2}-\frac{1}{\pi}\frac{\partial v}{\partial t}=0,\quad 0<\theta<\theta_0,\quad r>0,\quad t>0 \quad . \quad . \quad (1)$$

with

$$v=0,\quad \theta=0,\qquad r>0,\quad t>0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$v=1,\quad \theta=\theta_0,\qquad r>0,\quad t>0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and

$$v=0,\quad 0<\theta<\theta_0,\quad r>0,\quad t=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Multiplying (1) and its boundary conditions (2) and (3) by e^{-pt}, $p>0$, integrating with respect to t from 0 to ∞, and writing

$$\bar{v}(p)=\int_0^\infty e^{-pt}v(t)dt$$

for the Laplace transform* of v, we obtain the " subsidiary equation "

$$\frac{\partial^2 \bar{v}}{\partial r^2}+\frac{1}{r}\frac{\partial \bar{v}}{\partial r}+\frac{1}{r^2}\frac{\partial^2 \bar{v}}{\partial \theta^2}-q^2\bar{v}=0,\quad 0<\theta<\theta_0,\quad r>0, \quad . \quad . \quad . \quad (5)$$

where

$$q^2=p/\kappa. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

This has to be solved with boundary conditions

$$\bar{v}=0,\qquad \theta=0,\quad r>0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and

$$\bar{v}=1/p,\quad \theta=\theta_0,\quad r>0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

---

* For the Laplace transformation method in the form used here, *cf.* Carslaw and Jaeger, Phil. Mag. (7) xxvi. p. 473 (1938); Proc. London Math. Soc. (2) xlvi. p. 361 (1940).

We seek a solution of (5) of form

$$\bar{v}=\frac{\theta}{p\theta_0}+\sum_{n=1}^{\infty}\sin s\theta\int_0^\infty f_n(u)\mathrm{J}_s(ur)du, \quad . \quad . \quad . \quad . \quad . \qquad (9)$$

where

$$s=n\pi/\theta_0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (10)$$

The form (9) is chosen as consisting of the steady state solution together with a series of terms which satisfy the boundary conditions; *i. e.* (9) satisfies (7) and (8), and in order that it may satisfy (5) we require

$$\frac{\theta}{\kappa\theta_0}+\sum_{n=1}^{\infty}\sin s\theta\int_0^\infty (q^2+u^2)\mathrm{J}_s(ur)f_n(u)du=0,$$

or

$$\int_0^\infty (q^2+u^2)\mathrm{J}_s(ur)f_n(u)du=\frac{2}{\pi\kappa}\frac{(-)^n}{n}. \quad . \quad . \quad . \quad . \quad . \qquad (10)$$

Now

$$\int_0^\infty \frac{\mathrm{J}_s(ur)du}{u}=\frac{1}{s}.$$

So (10) will be satisfied if we take

$$f_n(u)=\frac{2(-)^n}{\kappa u\theta_0(q^2+u^2)}.$$

Thus, finally,

$$\bar{v}=\frac{\theta}{p\theta_0}+\sum_{n=1}^{\infty}\frac{2(-)^n}{\kappa\theta_0}\sin s\theta\int_0^\infty \frac{\mathrm{J}_s(ur)du}{u(q^2+u^2)}. \quad . \quad . \quad . \quad . \quad . \qquad (11)$$

Using the inversion theorem for the Laplace transformation we find from this

$$v=\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{\lambda t}\bar{v}(\lambda)d\lambda$$

$$=\frac{\theta}{\theta_0}+\frac{2}{\theta_0}\sum_{n=1}^{\infty}(-)^n\sin s\theta\int_0^\infty e^{-\kappa u^2 t}\frac{\mathrm{J}s(ur)du}{u}. \quad . \quad . \quad . \quad . \qquad (12)$$

It can be verified directly that (12) satisfies (1), (2), (3), and (4).

If *both faces of the wedge are kept at unity for* $t>0$, *and the initial temperature is zero*, we obtain

$$v=1-\frac{4}{\theta_0}\sum_{n=0}^{\infty}\sin\frac{(2n+1)\pi\theta}{\theta_0}\int_0^\infty e^{-\beta u^2}\frac{\mathrm{J}_s(u)du}{u}, \quad . \quad . \quad . \qquad (13)$$

where s takes the values $(2n+1)\pi\theta/\theta_0$ for $n=0, 1, 2, \ldots$, and $\beta=\kappa t/r^2$.

If $\theta_0=\pi$ the region becomes the region $z>0$, where $z=r\sin\theta$. Comparing the known solution for this region with that given by (13) we obtain the interesting expansion

$$\operatorname{erf}\frac{r\sin\theta}{2\sqrt{(\kappa t)}}=\frac{4}{\pi}\sum_{n=0}^{\infty}\sin(2n+1)\theta\int_0^\infty e^{-\kappa t u^2/r^2}\mathrm{J}_{2n+1}(u)\frac{du}{u}. \quad . \quad . \qquad (14)$$

3. *Numerical Calculations.*

The integrals appearing in (13) may be evaluated as infinite series, using the result*

$$\int_0^\infty e^{-\beta x^2}\frac{J_s(x)dx}{x}=\frac{\Gamma(\frac{1}{2}s)}{2^{s+1}\beta^{\frac{1}{2}s}\Gamma(s+1)}\,{}_1F_1\left(\tfrac{1}{2}s\,;\ s+1;\ -\frac{1}{4\beta}\right) \quad . . \quad (15)$$

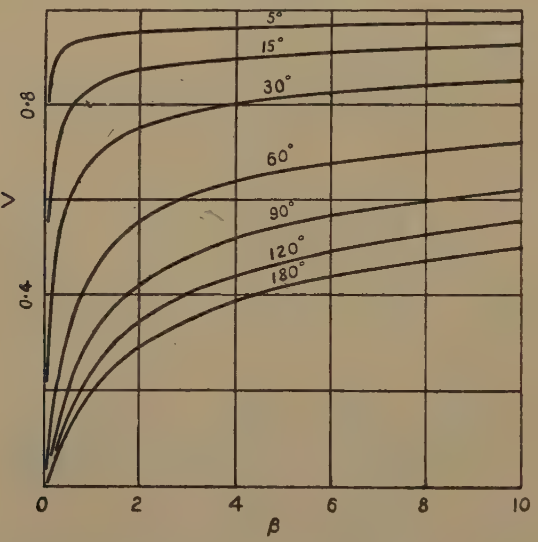
For small values of β they are evaluated most easily by using Watson, *loc. cit.*, § 3.2 (1), § 13.3 (5), and the asymptotic expansion for $I_\nu(x)$.

If $\beta>1$ the hypergeometric functions in (15) may be replaced by unity correct to three places of decimals.

Results are most simply given as values of v against $\beta=\kappa t/r^2$ for varying θ.

In fig. 1, curves are given for the wedge of angle 2π, the semi-infinite crack maintained at unit temperature, showing the variation of v with $\beta=\kappa t/r^2$ for the values 5°, 15°, 30°, 60°, 90°, 120°, 180° of θ_0. The corresponding results for the more important cases of the wedges of angles 270° and 90° are given in Tables I. and II. In these, as in the later tables, results have been calculated to four places, so it is hoped that the third figures given are correct.

Fig. 1.



4. *Heat Conduction in the region* $0<r<a$, $0<\theta<\theta_0$. *The initial temperature zero. The surface kept at unity for* $t>0$.

The solution of this problem may be derived by methods similar to those of § 2, but since it may also be obtained by the use of the Green's

---

* Watson, 'Theory of Bessel Functions' § 13·3 (2).

TABLE I.

θ β	5°.	10°.	15°.	30°.	45°.	60°.	90°.	120°.	135°.
0·1	·845	·698	·562	·262	·112	·050	·016	·010	·010
0·2	·890	·782	·680	·423	·253	·155	·073	·051	·049
0·3	·909	·819	·733	·508	·344	·236	·131	·098	·095
0·4	·920	·842	·765	·561	·404	·296	·180	·140	·135
0·5	·928	·856	·787	·599	·450	·341	·220	·175	·170
0·6	·933	·867	·803	·627	·484	·378	·253	·206	·200
0·7	·938	·876	·816	·649	·512	·408	·282	·233	·227
0·8	·941	·883	·826	·667	·535	·433	·307	·256	·250
0·9	·944	·889	·834	·682	·555	·454	·329	·277	·271
1·0	·946	·893	·841	·695	·571	·474	·348	·296	·290
2·0	·960	·919	·880	·767	·667	·584	·469	·416	·410
3·0	·965	·931	·897	·800	·712	·638	·532	·482	·476
4·0	·969	·938	·908	·820	·740	·672	·573	·525	·519
5·0	·971	·943	·915	·834	·760	·696	·602	·556	·551
6·0	·973	·947	·920	·844	·775	·714	·625	·581	·576
7·0	·975	·950	·925	·852	·786	·729	·644	·601	·596
8·0	·976	·952	·928	·859	·796	·741	·659	·617	·612
9·0	·977	·954	·931	·865	·804	·751	·672	·631	·626
10·0	·978	·956	·934	·870	·812	·760	·683	·644	·639

TABLE II.

θ β	5°.	10°.	15°.	20°.	30°.	40°.	45°.
0·1	·850	·706	·576	·464	·302	·224	·214
0·2	·903	·810	·723	·645	·527	·465	·458
0·3	·928	·859	·794	·736	·646	·598	·592
0·4	·943	·888	·836	·790	·717	·679	·674
0·5	·953	·907	·864	·825	·765	·733	·729
0·6	·960	·920	·884	·851	·799	·772	·768
0·7	·965	·931	·899	·870	·825	·800	·798
0·8	·969	·938	·910	·884	·844	·823	·820
0·9	·972	·945	·919	·896	·860	·841	·838
1·0	·974	·950	·927	·906	·873	·856	·853
2·0	·987	·974	·962	·951	·934	·925	·924
3·0	·991	·982	·974	·967	·955	·949	·948
4·0	·993	·987	·980	·975	·966	·962	·961
5·0	·995	·989	·984	·980	·973	·969	·969
6·0	·996	·991	·987	·983	·977	·974	·974
7·0	·996	·992	·989	·986	·981	·978	·978
8·0	·997	·993	·990	·987	·983	·981	·980
9·0	·997	·994	·991	·989	·985	·983	·982
10·0	·997	·995	·992	·990	·986	·984	·984

Function and the steady state solution* it will not be given in detail here.

The result is

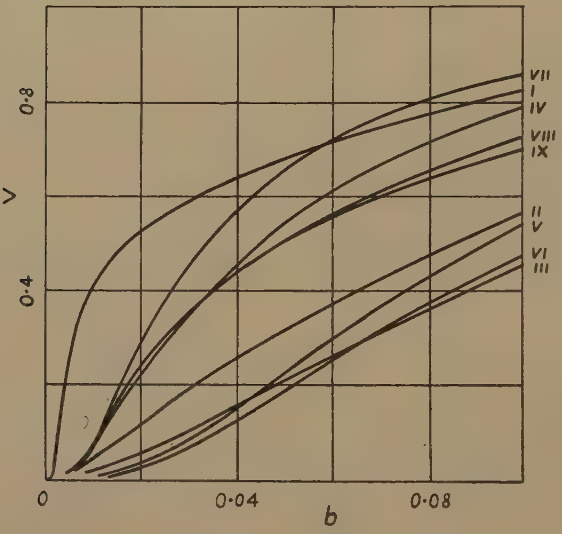
$$v=1-\frac{8}{\pi}\sum_{n=0}^{\infty}\frac{1}{2n+1}\sin\frac{(2n+1)\pi\theta}{\theta_0}\sum_{m=1}^{\infty}\frac{e^{-b\alpha^2_{s,m}}\,J_s(r\alpha_{s,m}/a)}{[J_s'(\alpha_{s,m})]^2}\int_0^1 tJ_s(\alpha_{s,m}t)dt, \quad . \quad . \quad . \quad (16)$$

where
$$b=\kappa t/a^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

s is written for $(2n+1)\pi/\theta_0$ in the orders of Bessel Functions, and $\alpha_{s,m}$, $m=1, 2, \ldots$ are the positive roots of

$$J_s(\alpha)=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

Fig. 2.



5. *Numerical Calculations from* (16).

The integrals occurring in (16) may be reduced by the formulæ

$$\int_0^x tJ_s(t)dt=s\int_0^x J_{s-1}(t)dt-xJ_{s-1}(x) \quad . \quad . \quad . \quad . \quad , \quad . \quad . \quad . \quad . \quad (19)$$

$$=\frac{s}{s-2}\int_0^x tJ_{s-2}(t)dt-\frac{2(s-1)x}{(s-2)}J_{s-1}(x). \quad . \quad . \quad . \quad . \quad (20)$$

If $\theta_0=\pi/r$, where r is an integer, they may be made to depend on † $\int_0^x J_0(t)dt$, and if $\theta_0=2\pi/r$, r odd, on Fresnel's integrals ‡.

---

* Carslaw, 'Conduction of Heat' (2nd Ed. 1921), §§ 92, 9. The result of § 2 cannot be obtained in this way from the Green's Function for the wedge.

† Watson, *loc. cit.* p. 752.

‡ Watson, *loc. cit.* Table V.

Since (16) involves the three parameters $b=\kappa t/a^2$, r/a, and θ, results as complete as those for the wedge cannot be given, but values at representative points in the interior sufficient for practical applications are given below.

In fig. 2 the values of v are plotted against $b=\kappa t/a^2$ for the sector of angle 2π, the solid cylinder with a single radial crack: Curves I., II., . . . IX. are for $r/a=\cdot 25$, $\theta=30°$, $90°$, $180°$; $r/a=\cdot 5$, $\theta=30°$, $90°$, $180°$; $r/a=\cdot 75$, $\theta=30°$, $90°$, $180°$, respectively. In Tables III. and IV. corresponding results are tabulated for the half and quarter cylinders $\theta_0=\pi$ and $\frac{1}{2}\pi$.

TABLE III.

$r/a=$	·25.	·25.	·25.	·5.	·5.	·5.	·75.	·75.	·75.
$\theta=$	30°.	60°.	90°.	30°.	60°.	90°.	30°.	60°.	90°.
$b=\cdot 01$	·376	·123	·077	·078	·002	·001	·097	·090	·090
·02	·532	·280	·211	·228	·048	·025	·300	·246	·246
·03	·613	·381	·312	·359	·136	·101	·460	·365	·360
·04	·669	·460	·393	·467	·235	·188	·577	·459	·446
·05	·714	·526	·464	·555	·331	·278	·661	·538	·518
·06	·753	·585	·529	·626	·419	·364	·724	·603	·579
·07	·786	·638	·588	·684	·496	·442	·773	·659	·632
·08	·815	·685	·640	·732	·564	·513	·811	·707	·679
·09	·840	·727	·687	·771	·623	·575	·841	·748	·721
·10	·862	·763	·728	·804	·674	·631	·866	·782	·757
·20	·969	·946	·938	·958	·927	·916	·972	·952	·944
·30	·993	·987	·985	·990	·983	·980	·993	·988	·987
·40	·998	·997	·997	·998	·996	·995	·998	·997	·997

TABLE IV.

$r/a=$	·25.	·25.	·25.	·5.	·5.	·5.	·75.	·75.	·75.
$\theta=$	15°.	30°.	45°.	15°.	30°.	45°.	15°.	30°.	45°.
$b=\cdot 01$	·679	·454	·377	·370	·094	·044	·254	·098	·092
·02	·804	·663	·612	·538	·254	·168	·517	·302	·261
·03	·861	·760	·723	·650	·412	·332	·664	·468	·413
·04	·896	·820	·792	·733	·544	·478	·755	·594	·542
·05	·920	·862	·840	·796	·649	·596	·814	·685	·640
·06	·939	·894	·878	·843	·729	·688	·861	·761	·726
·07	·953	·918	·906	·880	·792	·760	·894	·817	·789
·08	·964	·937	·928	·908	·840	·815	·919	·860	·838
·09	·972	·952	·944	·929	·877	·858	·938	·892	·876
·10	·979	·963	·957	·946	·906	·891	·952	·917	·904
·20	·998	·997	·997	·996	·993	·992	·997	·994	·993

6. *The Point of Slowest Heating.*

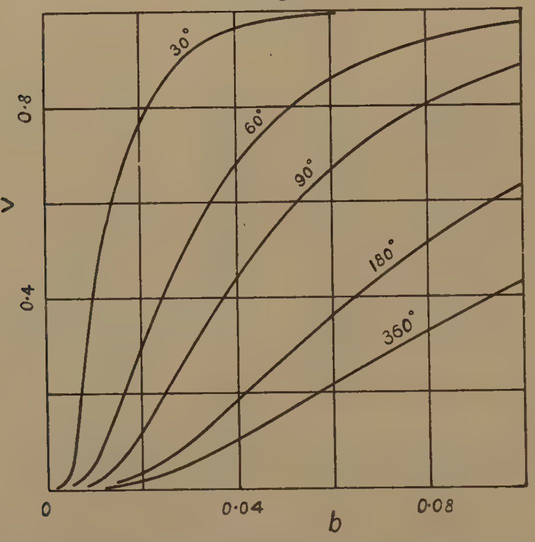
One of the most interesting physical problems is to determine what point of the sector $0<r<a$, $0<\theta<\theta_0$ will heat most slowly. Clearly this will be at $\theta=\frac{1}{2}\theta_0$, and at a radius chosen to make the coefficient of the largest exponential term in (16) as large as possible, *i. e.*, at the first maximum of $J_s(r\alpha_{s,1}/a)$, where $s=\pi\theta_0$ and $\alpha_{s,1}$ is the smallest positive root of $J_s(\alpha)=0$.

These values of r/a are given in Table V. for various values of θ_0, and in fig. 3 the heating curves for these points of slowest heating are shown, v being plotted against $b=\kappa t/a^2$.

TABLE V.

θ_0	360°	180°	90°	60°	30°
r/a	·371	·480	·595	·658	·755

Fig. 3.



7. *The Effect of a Difference of Radial and Tangential Conductivities.*

Suppose the conductivity in the direction of the radius is K_1 and that perpendicular to the radius is K_2, these being assumed constant. If the density and specific heat of the medium are ρ and c respectively, the equation of conduction of heat is

$$K_1\left[\frac{\partial^2 v}{\partial r^2}+\frac{1}{r}\frac{\partial v}{\partial r}\right]+\frac{K_2}{r^2}\frac{\partial^2 v}{\partial \theta^2}=\rho c\frac{\partial v}{\partial t}, \quad . \quad . \quad . \quad . \quad (21)$$

or, writing $\kappa_1=K_1/\rho c$ and $k^2=K_2/K_1$

$$\frac{\partial^2 v}{\partial r^2}+\frac{1}{r}\frac{\partial v}{\partial r}+\frac{k^2}{r^2}\frac{\partial^2 v}{\partial \theta^2}=\frac{1}{\kappa}\frac{\partial v}{\partial t} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

This differs from (1) only by the k^2 multiplying the term $\frac{1}{r^2}\frac{\partial^2 v}{\partial \theta^2}$, and the solution proceeds in precisely the same way save that in the orders of the Bessel functions $n\pi/\theta_0$ is to be replaced by $kn\pi/\theta_0$. Thus, for the wedge of angle θ_0, with zero initial temperature and surfaces kept at unity for $t>0$, the solution is

$$v=1-\frac{4k}{\theta_0}\sum_{n=0}^{\infty}\sin\frac{(2n+1)\pi\theta}{\theta_0}\int_0^{\infty}e^{-\kappa u^2 t}\frac{J_s(ur)du}{u},$$

where $s=(2n+1)\pi k/\theta_0$.

Thus, the solution for this case may be derived from the solution for $k=1$: the temperature at any time at the point (r, θ) in the wedge $0<\theta<\theta_0$ with $K_2/K_1=k^2$ is the same as that at the point $(r, \theta/k)$ in the wedge $0<\theta<\theta_0/k$ with $K_2/K_1=1$.

TABLE VI.

r/a b	0.	·1.	·2.	·3.	·4.	·5.	·6.	·7.	·8.	·9.
·01	·000	·000	·000	·000	·000	·001	·006	·041	·176	·506
·02	·000	·000	·000	·001	·004	·018	·059	·161	·356	·652
·03	·000	·001	·003	·008	·023	·059	·134	·266	·465	·722
·04	·004	·005	·011	·025	·055	·111	·206	·348	·539	·765
·05	·013	·016	·028	·051	·094	·164	·270	·414	·594	.796
·06	·029	·035	·051	·084	·137	·216	·326	·468	·636	·818
·07	·053	·060	·081	·120	·180	·264	·376	·513	·670	·836
·08	·082	·090	·115	·158	·222	·309	·420	·551	·698	·851
·09	·116	·124	·151	·197	·264	·351	·459	·584	·721	·863
·10	·152	·161	·189	·236	·303	·390	·494	·613	·742	·873
·20	·499	·506	·527	·561	·606	·662	·725	·794	·864	·934
·30	·718	·722	·734	·753	·779	·811	·846	·885	·924	·963
·40	·842	·844	·851	·861	·876	·894	·914	·935	·958	·979
·50	·911	·912	·916	·922	·931	·940	·952	·964	·976	·988
·60	·950	·951	·953	·956	·961	·967	·973	·980	·987	·994
·70	·972	·972	·974	·976	·978	·981	·985	·989	·993	·996
·80	·984	·985	·985	·986	·988	·989	·991	·994	·996	·998
·90	·991	·991	·992	·992	·993	·994	·995	·996	·998	·999
1·00	·995	·995	·995	·996	·996	·997	·997	·998	·999	·999

The same result holds for the region $0<\theta<\theta_0$, $0<r<a$ with conductivities K_1 and K_2 and constant surface temperature.

8. *Heat Conduction in the Infinite Cylinder, $r<a$. The surface kept at unity for $t>0$. The initial temperature zero.*

The solution of this problem is well known* to be

$$v=1+\frac{2}{a}\sum_{m=1}^{\infty}e^{-\kappa\alpha^2{}_m t}\frac{J_0(r\alpha_m)}{\alpha_m J_0'(a\alpha_m)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

the summation being taken over the positive roots of $J_0(a\alpha)=0$.

---

* Carslaw, 'Conduction of Heat (2nd Ed., 1921), § 110. For a solution using the Laplace Transformation, see Carslaw and Jaeger, *loc. cit.* The convergence of (23) is slow for small values of t, and approximations may easily be made as in Goldstein, Proc. London Math. Soc. (2) xxxiv. p. 51 (1932); see also Bromwich, *ibid.* (2) xxxi. p. 209 (1930.

In Table VI. a short table of v for values of r/a and $b=\kappa t/a^2$ is given.

In conclusion, I wish to thank Miss M. E. Clarke for her assistance with the computations.

The University of Tasmania.

LV. *The Space Relation among Five Points in Elliptic and Hyperbolic Geometry.*

By Ludwik Silberstein, Ph.D.*

[Received February 25, 1942.]

"Space relation" is the habitual name for the relation between the ten distances $r_{\iota\kappa}$ of a general pentad of points. The question as to the actual form of this relation in *Euclidean* space has attracted the attention of several eminent geometers. Just a century ago (Cambridge Math. Journal, 1841) Cayley gave this relation in a perfectly symmetrical form in which, that is, none of the five points is formally privileged, viz., by a vanishing determinant of the *sixth* order,

$$\begin{vmatrix} r_{11}^2 & r_{12}^2 & r_{13}^2 & r_{14}^2 & r_{15}^2 & 1 \\ r_{21}^2 & r_{22}^2 & r_{23}^2 & r_{24}^2 & r_{25}^2 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{51}^2 & r_{52}^2 & r_{53}^2 & r_{54}^2 & r_{55}^2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{vmatrix} = 0,$$

where, of course, all $r_{\kappa\kappa}^2$ are nil. The relation is thus homogeneous of the eighth degree in the distances. In 1889 (Proc. R. S. Edinburgh) Muir expressed the space relation by a vanishing determinant of the fourth order only (again of 8th degree in the $r_{\iota\kappa}$) in which, however, one of the five points is formally privileged †.

To the best of my knowledge the relation for a non-euclidean space has not been considered, and the object of the present note is to show that for a space of constant non-vanishing curvature, elliptic or hyperbolic, the contemplated metrical relation between five points can be derived most directly and easily from their purely projective, non-metrical, linear relation by merely applying to the latter the very operator

---

* Communicated by the Author.

† Namely, if r_1, r_2, r_3, r_4 be the distances of four points from the fifth, Muir's determinant consists of four rows such as

$$r_1^2,\ \tfrac{1}{2}[r_1^2+r_2^2-r_{12}^2],\quad \tfrac{1}{2}[r_1^2+r_3^2-r_{13}^2],\quad \tfrac{1}{2}[r_1^2+r_4^2-r_{14}^2].$$

(a polarity) which impresses metrics upon the manifold and thus endows every pair of points with a distance.

In fact, consider first a purely projective threefold. Let $P_1, \ldots P_5$ be any five of its points, in general no four coplanar. Then, using Möbius's point-calculus,

$$\alpha P_1+\beta P_2+\gamma P_3+\delta P_4+\epsilon P_5=0, \qquad \ldots\ldots \quad (1)$$

where α, β, etc. are mere numbers. This equation expresses only that the five points belong to the same projective three-space, namely, that P_5 (say) is on a line joining P_4 with any point of the plane $P_1P_2P_3$.

Now impress metrics upon the manifold, in the Cayleyan way, by means of an elliptic or a hyperbolic *polarity* (operator) Ω, so that ΩP will stand for the polar (plane) π of the point P and, at the same time, $\Omega\pi$ for the pole P of the plane π. If P and Q be any two points, $P\Omega Q$, the product of a point into a plane, is a pure number, such, moreover, that $Q\Omega P=P\Omega Q$ and vanishing only when P is on the polar of Q, and thus also Q on the polar of P.

The *distance* $r=\overline{PQ}$ of the two points * is then defined by

$$\operatorname{Cos} r=\frac{P\Omega Q}{\sqrt{P\Omega P\,.\,Q\Omega Q}},$$

where Cos stands for cos or cosh according as Ω is an elliptic or a hyperbolic polarity (Cayley's "absolute" $P\Omega P=0$ imaginary or real, respectively). In terms of (four) homogeneous projective point-co-ordinates, $P\Omega P$ is the familiar homogeneous quadratic function of the co-ordinates of P, and $P\Omega Q$ a homogeneous bilinear function of the co-ordinates of P and Q. It is, however, manifestly advantageous not to split our entities, the points, into co-ordinates, but to operate on these entities themselves.

It is convenient to normalize all points so that $P\Omega P=1$. (It will be recalled that, if $\lambda\neq0$ be any number, λP is congruent with P.) The general distance formula is thus simplified to $\operatorname{Cos} r=P\Omega Q$. In our case, for each of the ten distances between the five points,

$$c_{\iota\kappa}=\operatorname{Cos} r_{\iota\kappa}=P_\iota\Omega P_\kappa, \qquad \ldots\ldots\ldots \quad (2)$$

where $c_{\iota\kappa}=c_{\kappa\iota}$ will be used as a short symbol for $\operatorname{Cos} r_{\iota\kappa}$.

After these explanations, apply the operator Ω to (1). Then, since Ω is distributive,

$$\alpha\Omega P_1+\beta\Omega P_2+\ldots+\epsilon\Omega P_5=0.$$

Next, multiply by P_1 or $P_2, \ldots P_5$, remembering that such a "multiplication" is again distributive. Then, by (2) and since $P_\kappa\Omega P_\kappa=1$,

$$\begin{gathered}\alpha+\beta c_{12}+\gamma c_{13}+\delta c_{14}+\epsilon c_{15}=0\\ \ldots\ldots\ldots\ldots\\ \alpha c_{51}+\beta c_{52}+\gamma c_{53}+\delta c_{54}+\epsilon=0,\end{gathered}$$

which can be considered as five homogeneous linear equations for the ratios of the five numerical coefficients α, β, etc., and since these do not all vanish simultaneously, we have

---

* With the "space constant" or "curvature radius" taken as unit.

$$|c_{\iota\kappa}| = \begin{vmatrix} 1 & c_{12} & c_{13} & c_{14} & c_{15} \\ c_{21} & 1 & c_{23} & c_{24} & c_{25} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{51} & c_{52} & c_{53} & c_{54} & 1 \end{vmatrix} = 0. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

This is the required *space relation* among the five points. It is expressed by a symmetrical determinant of just the fifth order which contains only the mutual distance of the points through their cosines (ordinary or hyperbolic), and in which all five points are on an equal footing. Our determinant shares this feature with Cayley's, but its order is a unit lower, being equal to the number of points. The squares of the distances $r_{\iota\kappa}$ are replaced by their cosines $c_{\iota\kappa}$ which are even functions of $r_{\iota\kappa}$. In these cosines (with $\iota \neq \kappa$) our equation (3) is of the fourth degree. It is, at the same time, a quadratic for any one $c_{\iota\kappa}$ in terms of the remaining nine.

Such, then, is the space relation between five points in elliptic or in hyperbolic geometry.

The relation in Euclidean space ought to follow from (3) as a limiting case, for an infinite curvature radius or for all $r_{\iota\kappa}$ tending to zero, either in the elliptic or the hyperbolic space.

Take, *e. g.*, the hyperbolic case, when $c_{\iota\kappa} = \cosh r_{\iota\kappa}$, and put $r_{\iota\kappa} = a_{\iota\kappa} t$, where $a_{\iota\kappa}$ are finite constants and t an independent variable which we shall make tend to zero. Let $D(t)$ be our determinant $|c_{\iota\kappa}|$, considered as a function of t. Then, for $t=0$, $\mathrm{D}(0)=0$, since all its elements become equal 1. The first derivative is, with $s_{\iota\kappa} = \sinh r_{\iota\kappa}$,

$$\mathrm{D}' = \begin{vmatrix} 0 & a_{12}s_{12} & a_{13}s_{13} & \cdots \\ c_{21} & 1 & c_{23} & \cdots \\ \cdot & \cdot & \cdot & \cdot \end{vmatrix} + \begin{vmatrix} 1 & c_{12} \cdots \\ a_2 s_{21} & 0 \cdots \end{vmatrix} + \cdots .$$

Now, each of these determinants vanishes for $t=0$ (since each has full four rows of units and one of zeros). Thus $D'(0)=0$. Similarly, the reader will find that all derivatives up to $D^{\mathrm{vii}}(0)$ inclusively vanish, while $D^{\mathrm{viii}}(0)$ is the sum of five determinants of the fifth order such as

$$\begin{vmatrix} 1 & 1 & 1 & 1 & 1 \\ a_{21}{}^2 & 0 & a_{23}{}^2 & a_{24}{}^2 & a_{25}{}^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{51}{}^2 & a_5{}^{22} & a_{53}{}^2 & a_{54}{}^2 & 0 \end{vmatrix},$$

that is to say, of five determinants obtainable from $|a_{\iota\kappa}{}^2|$ by replacing its first, or second, etc., fifth row by units, and thus

$$\mathrm{D}^{\mathrm{viii}}(0) = \begin{vmatrix} 0 & a_{12}{}^2 \cdots\cdots & a_{15}{}^2 & 1 \\ \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & 1 \\ a_{51}{}^2 & a_{52}{}^2 \cdots\cdots & 0 & 1 \\ 1 & 1 \quad 1 \quad 1 & 1 & 0 \end{vmatrix}.$$

Hence, by Taylor's theorem, up to terms of higher order,

$$8\,!\,|c_{\iota\kappa}| = 8\,!\,D(t) = D^{\mathrm{viii}}(0)t^8 = \begin{vmatrix} 0 & r_{12}^2 & \cdots & \cdots & r_{15}^2 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ r_{51}^2 & r_{52}^2 & \cdots & \cdots & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{vmatrix},$$

which is Cayley's determinant. Its vanishing, then, expresses the space relation in Euclidean geometry.

Returning to the more general relation (3), we may illustrate it by applying it at least to one particular case. Let P_1, P_2, P_3, P_4 be the corners of a regular tetrahedron, in elliptic or hyperbolic space, and r_1, r_2, r_3, r_4 the distances of P_5 (any point of the space) from these four corners. Then $r_{12} = r_{23} = \ldots = r_{34} = e$, the edge of the tetrahedron. Thus, if we write $\operatorname{Cos} e = \eta$ and $\operatorname{Cos} r_\kappa = \rho_\kappa (\kappa = 1, 2, 3, 4)$, our determinant becomes

$$|c_{\iota\kappa}| = \begin{vmatrix} 1 & \eta & \eta & \eta & \rho_1 \\ \eta & 1 & \eta & \eta & \rho_2 \\ \eta & \eta & 1 & \eta & \rho_3 \\ \eta & \eta & \eta & 1 & \rho_4 \\ \rho_1 & \rho_2 & \rho_3 & \rho_4 & 1 \end{vmatrix},$$

which is easily developed into

$$|c_{\iota\kappa}| = (1-\eta)^2\{(1-\eta)(1+3\eta) - (1+2\eta)(\rho_1^2+\rho_2^2+\rho_3^2+\rho_4^2) + 2\eta(\rho_1\rho_2 + \ldots \rho_3\rho_4)\},$$

and since $\eta \neq 1$, we have, by (3),

$$(1+2\eta)(\rho_1^2+\rho_2^2+\rho_3^2+\rho_4^2) - 2\eta(\rho_1\rho_2 + \ldots + \rho_3\rho_4) - (1-\eta)(1+3\eta) = 0, \quad (4)$$

the relation between the four distances r_κ of any point of space from the four corners of a regular tetrahedron of given edge e *. In particular, when P_5 is the centre of the tetrahedron, $\rho_1 = \ldots = \rho_4 = \rho = \operatorname{Cos} r$, and (4) gives

$$4 \operatorname{Cos}^2 r = 1 + 3 \operatorname{Cos} e, \quad \ldots\ldots \quad (4\,a)$$

the relation between the edge e of the tetrahedron and the radius r of the circumscribed sphere, in hyperbolic or in elliptic space (Cos = cosh, cos). If the edge and the radius become infinitesimal fractions of the curvature radius of the space, then, up to terms of the fourth order, (4 *a*) reduces to $4 \pm \frac{3}{2}e^2 = 4(1 \pm r^2)$, in the hyperbolic and the elliptic cases respectively. Thus, in either case,

$$e = \frac{2\sqrt{2}}{\sqrt{3}}\, r,$$

the well-known relation for a Euclidean regular tetrahedron.

---

* These four distances, tied by the relation (4), can be used as the four coordinates of a point in hyperbolic or elliptic space.

If we develop the relation (4) itself for evanescent $r_1, \dots r_4$, e, then all terms up to those of the fourth order cancel mutually, and the relation between the distances of any point P of Euclidean space from the corners of a regular tetrahedron of fixed edge e becomes

$$\Sigma r_\kappa^4 - \tfrac{2}{3}\Sigma r_\iota^2 r_\kappa^2 - \tfrac{2}{3} e^2 \Sigma r_\kappa^2 + e^4 = 0, \quad \dots \dots \quad (4_0)$$

where, as in (4), the second sum contains each combination $\iota\kappa$ once only. If the point P moves into the centre of the tetrahedron, then $r_\kappa = r$, the first two terms cancel, and we have again $e^2 = \dfrac{8}{3} r^2$. If P coincides with one of the corners, one r_κ vanishes, the remaining three are all equal e, and the left-hand member of (4_0) vanishes identically. In fact, our original determinant $|c_{\iota\kappa}|$ vanishes identically, whenever two of the five points coincide. For there is, of course, no relation between the distances of a general tetrad of points in space.

For four *coplanar* points we have $\alpha P_1 + \beta P_2 + \gamma P_3 + \delta P_4 = 0$, whence, as before, follows the relation between their six distances $r_{\iota\kappa}$, in an elliptic or a hyperbolic plane,

$$\begin{vmatrix} 1 & c_{12} & c_{13} & c_{14} \\ c_{21} & 1 & c_{23} & c_{24} \\ c_{31} & c_{32} & 1 & c_{34} \\ c_{41} & c_{42} & c_{43} & 1 \end{vmatrix} = 0, \quad \dots \dots \quad (5)$$

with the previous meaning of the $c_{\iota\kappa}$.

Finally, for three *collinear* points, the relation becomes

$$\begin{vmatrix} 1 & c_{12} & c_{13} \\ c_{21} & 1 & c_{23} \\ c_{31} & c_{32} & 1 \end{vmatrix} = 0. \quad \dots \dots \quad (6)$$

or, in developed form,

$$1 - (c_{12}^2 + c_{23}^2 + c_{31}^2) + 2c_{12}c_{23}c_{31} = 0.$$

This may be considered as a quadratic for any one of the c's in terms of the remaining two. Thus, for instance,

$$c_{13} = c_{12}c_{23} \pm \sqrt{(1 - c_{12}^2)(1 - c_{23}^2)},$$

i. e., $\operatorname{Cos} r_{13} = \operatorname{Cos} r_{12} \operatorname{Cos} r_{23} \pm \operatorname{Sin} r_{12} \operatorname{Sin} r_{23}$, whence, in both the elliptic and the hyperbolic case, $r_{13} = r_{12} \pm r_{23}$, verifying the additivity of distances of collinear points.

Rochester, New York.
January 20, 1942.

LVI. *Comments on T. E. W. Schumann's*$^{(1)}$ *paper, " An Investigation concerning G. I. Taylor's Correlation Coefficient of Turbulence."*

By Chaim L. Pekeris, D.Sc.*.

[Received March 11, 1942.]

In the above-mentioned paper, Dr. Schumann questions the validity of a result obtained by G. I. Taylor$^{(2)}$ over twenty years ago, namely that the autocorrelation coefficient R(ξ) of a temporally homogeneous fluctuating quantity $p(t)$ has the development

$$\mathrm{R}(\xi)=1-\frac{\xi^2}{2\,!\,\overline{p^2}}\overline{\dot{p}^2}+\frac{\xi^4}{4\,!\,\overline{p^2}}\overline{\ddot{p}^2}-\ldots \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

He states : " As a matter of fact, in the case to be discussed below, it will be definitely demonstrated that the correlation coefficient is not an even function of the time, but that it assumes the form $e^{-k\xi}$. This result must be regarded as a matter of considerable importance, because Taylor's dictum concerning the form of R(ξ) has probably had a retarding influence upon the development of the theory of turbulence."

The case discussed by Dr. Schumann is one in which the probability F(p) that p shall be between p and $p+dp$ is governed by a Planck-Fokker equation of the form

$$\frac{\alpha^2}{k}\frac{\partial \mathrm{F}}{\partial t}=\alpha^2\frac{\partial^2 \mathrm{F}}{\partial p^2}+2p\frac{\partial \mathrm{F}}{\partial p}+2\mathrm{F}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

By obtaining a general solution of (2) and then particularizing it for a concentrated source, Dr. Schumann proves that

$$\mathrm{R}(\xi)=e^{-2k\xi/\alpha^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

1. We shall first remark that equ. (3) can be obtained from (2) without an explicit solution of the latter. Multiply (2) by p and integrate from $-\infty$ to $+\infty$, obtaining :

$$\frac{\alpha^2}{k}\int_{-\infty}^{\infty} p\,\frac{\partial \mathrm{F}(p)}{\partial t}\,dp=\frac{\alpha^2}{k}\frac{\partial}{\partial t}\int_{-\infty}^{\infty} p\mathrm{F}(p)\,dp$$

$$=\Big[\alpha^2(p\dot{\mathrm{F}}-\mathrm{F})+2p^2\mathrm{F}\Big]_{p=-\infty}^{p=+\infty}-2\int_{-\infty}^{\infty} p\mathrm{F}\,dp. \quad (4)$$

Now assume that $\overline{p^2}=\int_{-\infty}^{\infty} p^2\mathrm{F}(p)\,dp$ exists, ,then the integrated term vanishes and (4) reduces to

$$\frac{\alpha^2}{k}\frac{dp}{dt}=-2\overline{p}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where $$\overline{p}=\int_{-\infty}^{\infty} p\mathrm{F}(p)\,dp. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

---

* Communicated by the Author.

The solution of (5) is

$$\bar{p}(t+\xi)=\bar{p}(t)e^{-2k\xi/\alpha^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

If now (7) be multiplied by $\bar{p}(t)$ and averaged over t, we arrive directly at (3).

2. Secondly, we wish to state that the exponential correlation is by no means new, but that it can be found in Taylor's paper in which the development (1) is given. As to the bearing of the form (3) on (1), it is clear that, strictly, (3) should be written as

$$R(\xi)=e^{-2k|\xi|/\alpha^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

since, by definition, $R(\xi)$ cannot be greater than 1. The function in (8) does not possess a Taylor series development at $\xi=0$, but can be represented by a Fourier integral

$$R(\eta)=\frac{2}{\pi}\int_0^\infty \frac{\cos(\omega\eta)\,d\omega}{(1+\omega^2)}; \quad \eta=2k\xi/\alpha^2. \quad . \quad . \quad . \quad . \quad . \quad (9)$$

It is evident that any attempt to obtain a Taylor series for $R(\xi)$ by expanding the cosine term and integrating term by term, fails. The representation (9), furthermore, exhibits the physical reason for the impossibility of such a development. If $F(\omega)$ denote the spectral intensity of $p(t)$ at the frequency ω—a quantity corresponding to the sum of the squares of the sine and cosine transforms of a function of integrable square—then it was shown by Taylor(3) that

$$R(\eta)=\int_0^\infty F(\omega)\cos(\omega\eta)d\omega. \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Now the spectral intensity of $\dot{p}(t)$ is, by the above definition of $F(\omega)$, equal to $\omega^2F(\omega)$. The integral of this function with respect to ω gives $\overline{\dot{p}^2}$, and since, according to (9),

$$F(\omega)=\frac{2}{\pi}\frac{1}{(1+\omega^2)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

this integral diverges. Functions which have an exponential correlation coefficient have therefore no derivative and necessarily no derivatives of higher order. The reason why the correlation coefficient of such a function does not possess the development (1) is because the coefficients of the powers of ξ diverge. *Wherever these coefficients are finite, the development* (1) *holds.*

3. A process which has an exponential autocorrelation coefficient is known as a Markoff process. Such a process can be said to possess no *linear memory*, in the sense that a knowledge of a linear combination of the values of the function in the past cannot be used to improve the forecast of the function in the future, because the statistical behaviour of the function in the future is entirely determined by its present value.

The implication of Dr. Schumann that, after periodic terms have been eliminated, natural phenomena are Markoffian, is no more true than that all ordinary differential equations are of the first order. A simple illustration of a process which is not Markoffian is the fluctuating charge in a LRC series circuit, due to thermal agitation. This function has a derivative but no second derivative. Its correlation coefficient consists of a sum of two exponential terms, which may, however, be complex. Such a function can be said to possess a linear memory for *single* events in its past, but not for multiple events. The development (1) of its correlation coefficient stops with the ξ^2 term, the remainder being a function which can be represented only by a Fourier integral. It is clear that next in degree of intelligence is a function which possesses a linear memory for *two* events in the past, but not for a greater number of events. A knowledge of the values of the first and second derivatives of such a function in the present can be used to improve the forecast of a future value of the function. It is significant that, at least theoretically, one can determine the species of the function from its empirically obtained autocorrelation coefficient, without knowing the differential equations which govern it.

4. Finally, in the opinion of this writer, G. I. Taylor has no more impeded the development of the statistical theory of turbulence than did Maxwell the development of electromagnetic theory.

References.

(1) Phil Mag. xxxii. p. 471 (Dec. 1941).
(2) Proc. Lond. Math. Soc. xx. p. 196 (1920).
(3) Proc. Roy Soc. London, A, clxiv. p. 478 (1938).

LVII. *A Dynamical Treatment of the Elements of Heat.*

By G. Burniston Brown, M.Sc., Ph.D., University College, London *.

[Received April 20, 1942.]

Abstract.

The elements of the subject of heat are developed from consideration of the dynamical effects produced by heat in a similar manner to that used in the development of electric and magnetic theory. The advantages are that a dynamical scale of temperature independent of the properties of any particular substance follows naturally and clearly, and that the unit of quantity of heat becomes the erg, also independent of the properties of any particular substance. This leads to a foundation of calorimetry directly on the principle of the conservation of energy instead of, as at present, on a usually undisclosed historical background of an indestructible caloric fluid. The relation between heat measured

---

* Communicated by the Author.

in calorimetry and heat measured in dynamical terms is obtained without recourse to Carnot's cycle and the obscurities associated with perfectly reversible engines. If a practical unit of temperature is taken to be 10^7 greater than the dynamical unit proposed, the specific heat of water at 15° C. becomes 1·007 joules per gram per degree. Knowledge of a gas constant and a mechanical equivalent of heat is not required in the system proposed.

> "But whatever may be the range of mechanical theories, they do not apply to the effects of heat. These make up a special order of phenomena, which cannot be explained by the principles of motion and equilibrium."—J. B. J. Fourier (1822). 'The Analytical Theory of Heat.' Cambridge University Press.

§ 1. *Introduction.*

It is not the object of this paper to trace the causes which led to the view that mechanical principles could not be applied in the study of heat. The object is to approach the subject from a viewpoint founded on a different philosophy of science. Instead of looking upon the phenomena of heat as a "special order" requiring therefore special units and special dimensions, we shall start by considering the division of physics into different subjects as having arisen in general from the division of the nervous system of the human body into different sense organs. An electric spark, for instance, which in nature is one, can be divided into all the special orders, heat, light, electricity, magnetism, and properties of matter. These differing bodily sensations are the cause of our ideas of the differing "nature" of the phenomena. Nothing can alter the fact that a charged piece of matter can produce different sensations in our bodies from an uncharged piece of matter, and we rightly say that they are different in their "nature." This is a fact which must be included in science and, indeed, Cavendish made use of such electrical sensations to turn himself into a galvanometer.

But when we come to *exact* science, that is to say, to that part of science which is concerned with the introduction of numbers into essentially-non-numerical nature, then we may well ask whether the units and dimensions so introduced have any connexion with the "nature" of the physical quantity concerned.

If, for instance, e represents a charged particle and m an uncharged particle, then they do symbolize particles of different nature. e. g., e can be produced by rubbing one's sleeve with a fountain pen, and disappear again as easily : this is not true of m. But if e stands for, say, 10 e.s.u. and m for 10 grams, then they now *symbolize numbers produced by conventional measuring operations.* They are, as Eddington says, measures of the mutual interactions of two particles. We have only to remember that this mutual interaction is, in principle, measured by the same type of operation for both e and m, e. g. a torsion balance, and that the formulæ involved are identical in form, to realize that the units and dimensions obtained by making the constant of proportionality unity and dimensionless, cannot symbolize the difference in "nature" between charged and uncharged particles.

When, therefore, the question arises of making the study of heat an *exact* science, that is to say, when we consider the introduction of numbers to measure the effects of heat, we do not have to concern ourselves with the nature of heat. On the contrary, believing that all nature is one and that divisions are artificial, we should attempt to introduce numbers *consistently* into all branches of physics.

If we introduce numbers into mechanics by using Gauss's conventional definition for the measurement of Newtonian force, then we should do the same in heat, light, sound, electricity, and magnetism. Thus in each case we should seek some change in the phenomena observed which is associated with a measurable *force*. In electricity and magnetism we use the measurement of the force of attraction and repulsion in order to introduce dynamics and define the units. In the case of heat we must ask: what changes due to heat are associated with force? The obvious answer is *expansion*; for expansion is an observable change due to heat and it is associated with force. The sensation of heat is an observable change but is not readily associated with force. The force of expansion most suitable for measurement is that of a gas.

The change in philosophy of science which has been indicated above may be summed up by a comparison. Instead of saying, as Helmholtz did in a famous lecture in 1869, that "the final aim of Natural Science is to discover the motions underlying all alterations and the motive forces thereof; that is, to resolve itself into Mechanics," we say "Exact science deals with the introduction of numbers into nature. For the sake of consistency and to avoid unnecessary confusion, we should introduce them into all branches of science according to the *same system*. The system of measurement of mechanics was historically the first and, in addition, has been employed in founding electricity and magnetism, and electromagnetism on a dynamical basis (*i. e.*, on measurement of force). The aim, therefore, of exact science should be to found all measurements on those of mechanics."

§ 2. *Temperature Scale.*

Without any scale for measuring temperature we can use our sensation of heat to convince ourselves that the volume, say, of a gas at constant pressure expands as it gets hotter and contracts as it gets colder. We then say that it expands with increase in temperature and *vice versa*. We can therefore define a system at constant temperature as one which is in thermal equilibrium with a gas at constant pressure whose volume remains constant, and this will, of course, agree with the cruder sensation of heat. Experiment then shows that, at constant temperature, the behaviour of gases can be expressed by Boyle's law, viz.,

$$pv \propto \mathrm{M},$$

where p is the pressure, v the volume, and M is the mass of gas having the pressure and volume given. Now, as has been pointed out, the sensation of heat alone is sufficient to inform us that, if the temperature does not remain constant, the product pv increases with increase in temperature.

Let us imagine, therefore, a gas which obeys Boyle's Law exactly at all temperatures, and define the measurement of temperature by writing

$$pv=k\mathrm{M}\theta.$$

This procedure is arbitrary, of course, like all definitions, but no more so than the use of the expansion of the mercury of a thermometer or the definition of equal time intervals, and it achieves our aim of defining temperature entirely in terms of mechanical quantities.

The units and dimensions of temperature are obtained in the correct way, viz., by choosing them so as to make the constant k both unity and dimensionless, so that it can be dropped entirely from the equation. We then have for our fundamental equation

$$pv=\mathrm{M}\theta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

A physical system will therefore be at a temperature of unity when it is in thermal equilibrium with a gram of the chosen ideal gas having a volume of 1 cm.^{3} and exerting a pressure of 1 dyne per cm.^{2}.

The dimensions of temperature are correspondingly

$$[\text{temperature}]=\left[\frac{\text{pressure}\times\text{volume}}{\text{mass}}\right]=[\text{velocity}^2]$$

agreeing with the result of a former treatment of dimensions *.

A temperature of double the unit temperature will be that of a body in thermal equilibrium with a gram of the chosen ideal gas for which the product pv is two. Similarly, a body will be said to have a temperature n when it is in thermal equilibrium with a gram of the ideal gas for which the product pv is n. Since the product pv is a measure of energy, the temperature scale is defined in terms of equal increments of energy.

Thus we can build up a temperature scale which is obviously a dynamical scale, and which is independent of the properties of any particular gas, since the thermometric substance is an ideal one. It is therefore an absolute dynamical scale of temperature. Before this scale can be utilized, we must examine how far any of the real gases deviate from Boyle's law, and determine the deviation, if any, as accurately as possible. Also the size of the degree, or magnitude of the scale unit, is not determined by (1) until we specify a particular gas. There are several reasons for the choice of hydrogen to define the degree of temperature: firstly, it obeys Boyle's law very closely; secondly, in actual practice it is convenient; and thirdly, it can be used over a greater range than other gases (with the exception of helium). Adopting hydrogen, therefore,

---

* "A New Treatment of the Theory of Dimensions," Proc. Phys. Soc. liii. p. 418 (1941).

for this purpose, we may proceed to calculate the relation between the Centigrade degree and the dynamical degree which will be written $°\Theta$.

We have $pv = \mathrm{RT}$ for a gram molecule.

When $p = 1$ dyne/cm.^{2} and $v = 1$ cm.^{3} T°A will be $1°\Theta$, so for hydrogen

$$1 \times 1 = \frac{83{\cdot}14 \times 10^6 \mathrm{T}°}{2{\cdot}016}$$

or

$$\mathrm{T}° = 1°\Theta = 2{\cdot}424 \times 10^{-8}\,°\mathrm{A}.$$

Like other units defined from the dyne, this is inconveniently small, and in practice a unit 10^7 times greater might be taken, giving a practical unit of temperature of approximately 0·24 deg. C. The zero of the scale so constructed would, of course, be the same as that of the absolute Centigrade scale (−273·2° C.), the temperature at which the product pv is zero. Thus, for example, we should have (writing deg. B for the practical unit)

$$0°\ \mathrm{C.} = 273{\cdot}2°\ \mathrm{A} = 1127°\ \mathrm{B}.$$

$$100°\ \mathrm{C} = 373{\cdot}2°\ \mathrm{A} = 1540°\ \mathrm{B}.$$

§ 3. *Quantity of Heat.*

In the attempt to introduce dynamical concepts and methods of measurement into the study of heat, we have looked for some manifestation of heat which involved force. In this way we have used the conventional measure of force, which is the foundation of mechanics, in order to measure temperature or degree of heat. Continuing, now, with our general viewpoint that the subjects of physics are arbitrary divisions, and that nature is really a unity, we naturally expect that the concept of energy, which is of such great importance in mechanics, will also be of great value in heat. We search, therefore, for a means of introducing this concept and its measure into heat. Having already discovered and measured a force due to heat, we have, of course, only to let it move its point of application through a known distance in order to have a given amount of work done, and we notice at once that the left-hand side of the fundamental equation

$$pv = \mathrm{M}\theta$$

has the dimensions of energy, so that the product $\mathrm{M}\theta$ has also the dimensions of energy.

Now, as has been already noticed, the measurement of p and v is really only practicable in the case of gases, but the product of the mass and the temperature ought to be measurable in a far wider range of phenomena. Thus if we consider, as in the case of energy, an isolated system, we expect to find that the sum of the products

$$\mathrm{M}_1\theta_1 + \mathrm{M}_2\theta_2 + \mathrm{M}_3\theta_3 + \ldots \text{ etc.},$$

will remain constant during any heat interaction, provided that any other interactions that can be measured in terms of energy, such as chemical action, mechanical work, changes of state, etc., do not occur. Experiment shows that this is closely true provided that $\mathrm{M}_1\ \mathrm{M}_2\ \mathrm{M}_3$ etc.

refer to different masses of the same substance in the same state, and that the temperature range is not too large. If M_1 M_2 M_3 etc. stand for masses of different substances, then experiment shows that

$$s_1M_1\theta_1+s_2M_2\theta_2+s_3M_3\theta_3+\ \dots\ =\text{constant},$$

where s_1 s_2 s_3 etc. are dimensionless factors which are different for each substance but sensibly constant provided the temperature variation is not too large.

It is not necessary here to recount the experiments which lead to the idea of *quantity of heat.* It is enough to recall that they lead to the view that quantity of heat depends on

(*a*) the substance with which it is associated;
(*b*) the mass of the substance;
(*c*) the temperature of the substance.

The dynamical concept of energy is therefore capable of representing the idea of *quantity of heat,* and its measure will act as a measure of quantity of heat in a body in the form $sM\theta$, where s is called the *specific heat* and defined as the energy (ergs) required to raise the temperature of unit mass of the substance through one degree in a given part of the temperature scale. We shall also have justification in calorimetry for assuming that the quantity of heat lost by the hot body will be equal to that gained by the cold body, since this will be an application of the principle of the conservation of energy, already known to be a metrical characteristic of nature.

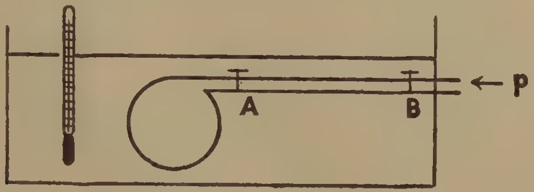
It is well known, however, that this measure-convention $sM\theta$ is not sufficient to accord entirely with the idea of quantity of heat: this is due, of course, to the experimental fact that heat measured in this way (*e. g.*, fall in temperature of water in a calorimeter) can "disappear" into another substance (*e. g.*, ice) without any corresponding rise of temperature. The measure-convention $sM\theta$ gives zero for the quantity of heat absorbed by the ice. We can get over this difficulty in a similar way to that by which we get over it in mechanics. We invent the concept of *potential* energy or *latent* heat, and just as, in mechanics, we associate it with position (since to associate it with velocity would give zero energy), so in heat we associate it with some other observable change —"change of state"—(since to associate it with temperature change would give zero heat). By defining the latent heat as the energy required to change the state of one gram of the substance at a given constant temperature, we ensure that the heat-energy equation balances and that the principle of the conservation of energy and the practice of calorimetry remain valid. This involves, of course, the reversibility of the heat-change-of-state relationship: this assumption is justified by experiment.

It follows from the arguments given above that, as regards a unit in a dynamical treatment of heat, we should choose the erg.

If the practical unit of temperature suggested in §2 is adopted, then the specific heat of water at 15° C. turns out very conveniently to be

1·007 joules per gram per deg. B. Thus, when an accuracy of not more than 1 per cent. is required, the specific heat of water can be taken as unity. For higher degrees of accuracy, the variation of specific heat with temperature must be taken into account.

It remains now to show how the work done by the force of expansion of a gas may be used to obtain a measure of a quantity of heat, and how the two systems of measurement (pv and $sM\theta$) can, in principle, be related in a direct and simple manner. For this purpose we can consider a simple system which attains the same final state by two different operations.



Suppose that we have a mass of any real gas in a conducting bulb with a straight non-conducting tubular portion, placed inside an insulated calorimeter and surrounded by water. The tubular portion has two non-conducting taps, A and B, and the end is open to the atmosphere whose constant pressure, equal to that of the gas inside, is p. We may suppose that we have a mercury thermometer already calibrated in practical dynamical degrees (deg. B), and that everything is at a uniform temperature t_1 (well above the liquefying point of the gas). Let AB be evacuated. Now perform the two following operations, each of which leads to an identical final state.

(1) Turn tap A and allow the gas to expand slowly into the vacuum from volume v_1 to volume v_2. Add hot water until the pressure returns to p. Let M grams of water at temperature t_3 be thus added and the final temperature t_2.

(2) Starting from the same initial conditions as in (1) but with a frictionless non-conducting piston at A and the tap B open to the atmosphere, add hot water and allow the gas to expand slowly against constant pressure p to volume v_2. Suppose this time a mass M of water at temperature t_4 has to be added, and that the final temperature is again t_2.

Since the final state of everything is the same in (1) and (2) (bulb, water, gas, calorimeter, etc.), the measure of the only observable change in the cause must be the equivalent of the measure of the only observable change in the effect, *i. e.*, the extra quantity of heat measured in (2) must be the equivalent of the external work done, measured by the product of the pressure and the volume change. So we have, by the principle of conservation of energy,

$$sM(t_4-t_2)-sM(t_3-t_2)=p(v_2-v_1)$$

or

$$sM(t_4-t_3)=p(v_2-v_1).$$

Therefore the quantity of heat given out by a mass M of water in cooling from t_4 to t_3 is $p\ (v_2-v_1)$ ergs.

We have thus measured a quantity of heat directly in ergs. This, of course, is in general inconvenient, and so this equation may be used to find the specific heat of water :

Mean specific heat of water between t_3 and t_4,

$$s=\frac{p(v_2-v_1)}{M(t_4-t_3)}.$$

In this direct way we can, in principle, find the value of the specific heat of water in joules per gram per deg. B. We can then employ this result to find the specific heats and latent heats of other substances according to already established procedure.

The practical unit of heat might therefore be taken to be the joule and would be the heat required to raise the temperature of one gram of water 1·007° B at room temperature (15° C. or 1188° B).

§ 4. *Calculation of the Mechanical Equivalent.*

On the view advanced in this paper the unit of heat is taken naturally to be the erg. No question, therefore, of the value of the mechanical equivalent arises, any more than would the question of the mechanical equivalent of electricity.

It follows from this that we should be able to calculate the mechanical equivalent of the calorie without resorting to direct measurement, just as we calculate the energy dissipated in an electric circuit.

To do this we can use the experiment described above for the case of hydrogen. Suppose that we have a mole of hydrogen in the bulb and that we raise the temperature 1° C. in operations (1) and (2). The heat supplied in (2) is, of course, C_p. If the expansion of volume is made small compared with the total volume of the gas, then the quantity of heat associated with the Joule-Thomson effect will be negligible, and we can call the heat supplied in (1) C_v.

Therefore for 1 mole of hydrogen raised from 15° C. to 16° C.,

$$(C_p-C_v)=\frac{p(v_2-v_1)}{J}=\frac{R}{J},\ \text{whence } J=\frac{R}{(C_p-C_v)}.$$

Latest values give

$R=8{\cdot}314\times10^7$ ergs per deg. C. per mole.

$C_p=6{\cdot}86$ calories per mole.

$C_v=4{\cdot}87$ calories per mole.

Thus $$J=\frac{8{\cdot}314\times10^7}{1{\cdot}99}=4{\cdot}18\times10^7 \text{ ergs per calorie.}$$

This method of calculating the mechanical equivalent is at least a hundred years old, but the object of presenting it again in somewhat modified form and with the latest available figures, is to emphasize the point made above that there should be no need to measure J directly.

It is not merely, as Saha calls it, "an ingenious method"; it is a fundamental and perfectly legitimate way of obtaining the value of the mechanical equivalent.

§ 5. *Conclusion.*

A dynamical treatment of the elements of heat has been suggested which leads directly to a dynamical temperature scale, which, since it applies to an imaginary (ideal) gas, is necessarily independent of the properties of any particular real gas and therefore absolute. The measurement of quantity of heat in ergs is suggested in a simple manner by the fundamental equation defining the temperature, and the methods of calorimetry are founded directly on the principle of conservation of energy.

The simple thermodynamical operation described enables the relation between the calorimetric measure of a quantity of heat and the dynamical measure of the equivalent quantity of work to be found without recourse to reversible cycles, infinite thermal capacities, heat flow between bodies at the same temperature, or other obscurities of Carnot's Cycle.

The thermodynamical scale proposed, being an absolute dynamical scale, is therefore the equivalent of Lord Kelvin's absolute scale, although much more simply derived.

§ 6. *Acknowledgements.*

The author's thanks are due to Mr. Orson Wood for some helpful criticism and to Mr. J. H. Awbery for the latest values of the physical constants used.

LVIII. *Experiments with Helmholtz Resonators.*

By H. Mary Browning, M.Sc., Ph.D., F.Inst.P.,
University College, Nottingham *.

[Received April 27, 1942.]

1. *Introduction.*

According to Helmholtz theory the fundamental frequency of vibration of a resonator varies inversely as the square root of its volume. A series of experiments has been carried out to determine under what conditions and to what accuracy the law is fulfilled in the case of a cylindrical resonator with a neck of constant length.

---

* Communicated by the Author.

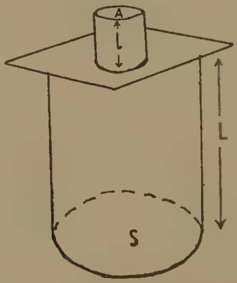
2. *Theory.*

The shape of the resonator was as shown in the diagram, fig. 1; l is the length of the neck, L that of the body, A the area of cross-section of the neck, S that of the body. N is the frequency of the note which produces resonance and c the velocity of sound taken in free air. Writing $k=2\pi N/c$ and $K=A/l$ (the conductivity of the neck), Richardson (1928) showed that $\tan kL=K/kS$.

3. *Experiments.*

Brass necks of 2·35 cm. length and varying cross-section were soldered on thick tin bases about 12 cm. square. Brass tubes of 1·1 mm. thickness and varying diameters were fastened in turn by plasticine symmetrically to the tin. These were lowered into water and a sounding tuning-fork held about 1 cm. above the neck; the tube was moved about until the position of resonance was obtained. The length from the top of the water to the underside of the tin plate was measured. Six values of L for each fork and tube were taken and the mean calculated.

Fig. 1.



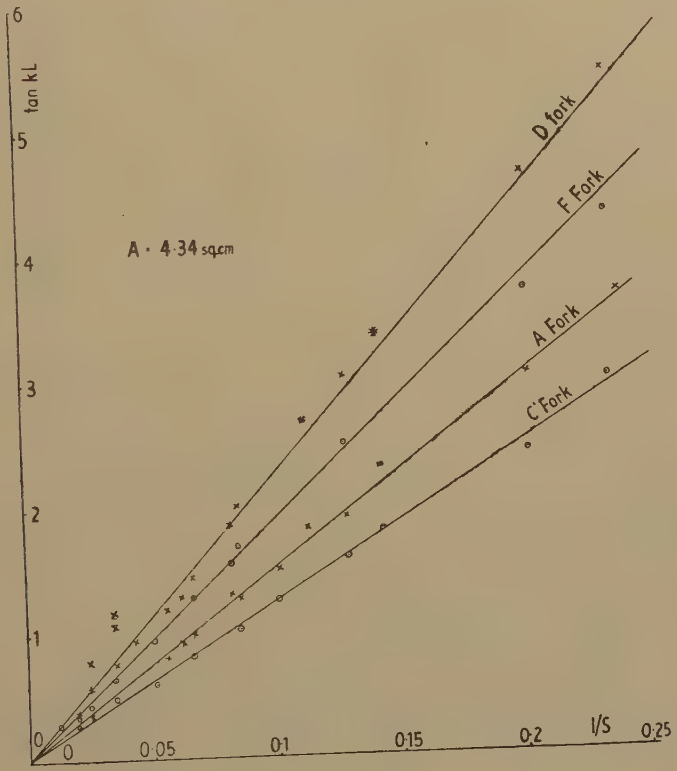
From the velocity of sound, c, obtained for the temperature of the experiment, the value k was calculated; $\tan kL$ was then plotted against $1/S$ (fig. 2) where S is the cross-section of the body of the resonator. In making these experiments, it was assumed that the frequency N of the fork and the dimensions of the resonator were constant, since the maximum variation in temperature was less than 6° C. It should be noted, however, that l/kS is not independent of temperature since N is assumed to be constant. Variation in temperature accounts for some deviation of the points from the straight line. The velocity of sound is affected by the narrowness of the tube, but the results show little discrepancy due to taking the value of c for free air.

It was thought that the humidity of the air above the water might affect the results, so a set of tubes was used in a cylinder containing olive oil; however, since no appreciable difference from the experiments with water was noticed, the remainder were all taken using water.

Four forks marked D (288·0), F (341·3), A (426·6) and C (512) were taken, and the tubes used as the body of the resonator were varied from the same diameter as the neck to about 9 cm. Four sizes of neck were chosen, 1·57 cm., 2·35 cm., 2·98 cm., and 3·64 cm. in diameter, and all were 2·35 cm. long.

When the diameter of the body was equal to that of the neck, the latter with its flange was attached and the measurements of L, the length of the body, taken from the water to the flange. If a tube of uniform cross-section from the water to the top has its total length measured as L, there is a discontinuity in the value of tan kL as compared with values for larger cross-sections of the body.

Fig. 2.



4. *Results.*

1. Graphs of fig. 2, as of three similar figures not reproduced, show that tan kL is proportional to 1/S. It is noticed that at a few places the points deviate rather far from the straight line, but there are three reasons which may account for this; (*a*) the wall of the tube was sometimes thinner (readings on the graph below the mark *) or thicker (*) than the average, which was 1·1 mm., (*b*) the tube was irregular (⁑) and (*c*) when 1/S was large, the errors were exaggerated, as the length of the body of the resonator was as small as 2 cm. in one case and a 5 per cent. error in measurement was possible.

2. From the slopes of the graphs of fig. 2 (and three similar graphs), the appropriate values of S tan kL were obtained. From these values were calculated corresponding ones of L for a series of values of S: the length L of the body of the tube was multiplied by the cross-sectional area to obtain the volume, and graphs were plotted between the volume V and the cross-section S for the four forks marked D, F, A and C, for each value of the cross-section A of the neck. The relevant graphs are represented by fig. 3.

Fig. 3.

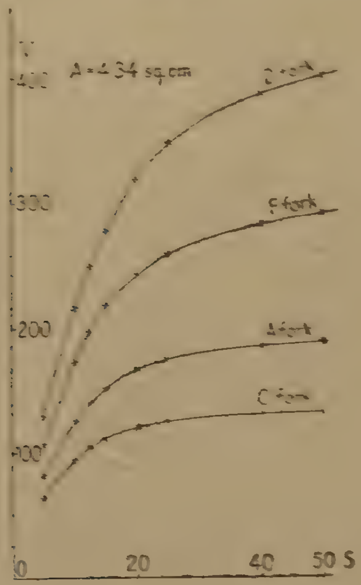
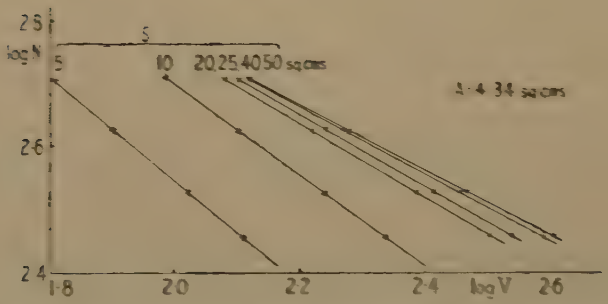


Fig. 4.

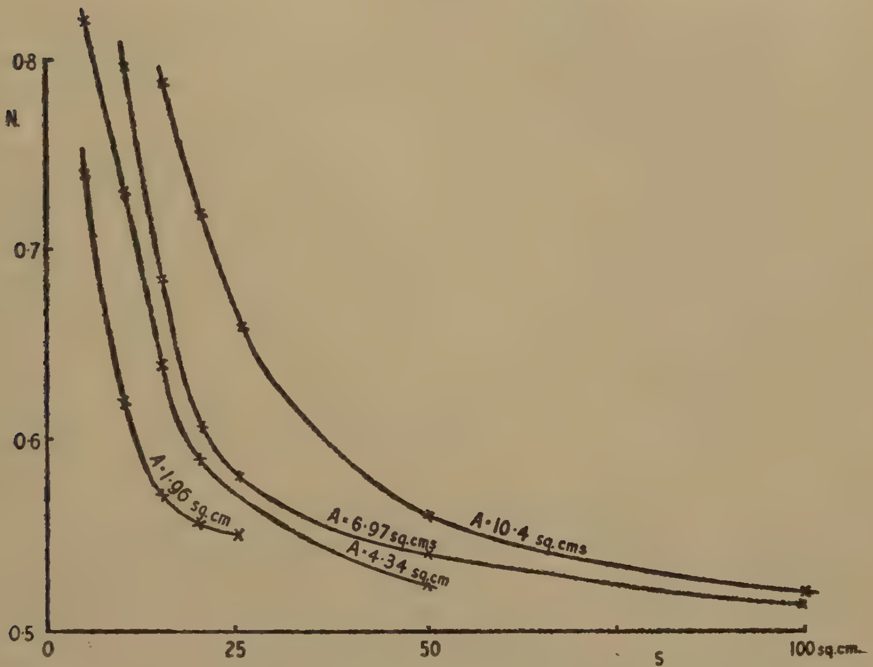


3. Taking particular values of the cross-section S of the body of the tube, *e.g.* S=50 sq. cm., the data of fig. 4, and others, were obtained showing the relation between log V and log N, where V is the volume of the inside of the body and N is the frequency of the vibration. Since

the graphs were linear, $N \propto l/V^n$, where the value of n varies according to the width of the neck. The manner in which n varies with S can be seen from fig. 5.

4. Substituting for k in equation 1, $\tan kL = (A/kl)1/S$, we have $Alk = \tan \frac{kL}{1/S}$, given by the slopes of the graphs in fig. 4; the value of l was calculated, assuming A was the actual measured area of the neck

Fig. 5.



of the tube. These values are given below, together with the end correction to be added to the diameter of the neck.

Area of Neck	1·94	4·34	6·97	10·4 sq. cm.
l for D fork..........	3·24	3·65	3·78	4·034 cm.
l for F ,, ...	3·18	3·63	3·82	4·004 ,,
l for A ,, ...	3·22	3·64	3·75	4·065 ,,
l for C ,, ...	3·14	3·62	3·92	4·105 ,,
Mean l	3·195±0·05	3·64±0·03	3·82±0·1	4·07±0·11 cm.
Diam of Neck	1·57	2·35	2·98	3·56 cm.
End Correction	0·75±0·05	1·29±0·03	1·57±0·1	1·72±0·11 cm.

The graph of the end-correction and diameter of neck gave a straight line, the slope of which was 0·5. Hence the end-correction is half the diameter of the neck.

5. *Conclusions.*

1. Using cylindrical resonators with cylindrical necks of constant length but varying diameter, the formula $\tan kL = K/kS$ is verified in so far as a linear graph is obtained between $\tan kL$ and $1/S$; but the value found for K=Area of neck / its length is not the measured value. The area is assumed correct. (Fig. 2.)

2. The end-correction to be added to the length of the neck varies approximately linearly with the diameter of the neck. For a neck of actual length 2·35 cm. the correction is 0·5 diameter.

3. Fig. 3 shows that, for a particular frequency, the volume of the body of the resonator does not remain constant as its area changes but varies considerably, approaching a constant value as the area becomes greater; thus the frequency of a Helmholtz resonator is not constant for constant volume, but depends on the dimensions.

4. A formula approximating to $N \propto 1/V^n$ is obtained between the frequency N and the internal volume V of the body of the resonator, where N varies between 0·5 and 0·8. A graph (fig. 5) shows how n varies with V for different values of the frequency N, and it is found that for high values of N and small diameters of the necks, n approaches 0·5, the value which agrees with Helmholtz's theory.

Reference.

Richardson, E. G., Phys. Soc. Proc. xl. p. 206 (1928).

LIX. *Notices respecting New Books.*

The Birth and Death of the Sun. By GEORGE GAMOW. [Pp. xv+238.] (London: Macmillan & Co., 1941. Price 12*s.* 6*d.*)

BY the quantitative demonstration that the stream of energy from a star may be generated by nuclear transformations occurring in its interior, a long-standing problem has been carried much nearer to solution. There has been a long and arduous climb over difficult terrain, and it is fitting that one who has blazed the trail should pause to take breath and survey the scene as a whole from the new vantage point gained. So here Gamow gives "in the simplest terms of which he is capable, an outline of the fundamental discoveries and theories that now permit us a general view of the evolution of our world."

After introducing the central problem of the sun and its energy, the author describes the essentials of the relevant knowledge of atomic and nuclear properties. He discusses thermonuclear processes, and gives a full account of the manner in which the transformation of hydrogen into helium may occur in association with a cyclic process involving carbon and nitrogen. This cycle would seem to be of predominant importance in the main sequence stars including the sun. For the earlier stars, such as the red giants, non-cyclic transformations can be plausibly assigned. The later evolutionary stages are associated with contraction of the assembly of fully-ionized atoms towards the single massive nucleus constituting the dead star. The end may be foreseen, but the sun may be a larger and much more luminous star before the final collapse sets in. Beginnings are much more speculative, but Gamow does not hesitate to go back to a time when an unexpanded universe had a density of several billions, and to link this up with the conditions necessary for the formation of the heavy radioactive elements.

Enlivening, but not disturbing, the main argument, discussions are given of all sorts of related topics such as the harnessing of atomic energy, pulsating stars, novæ, and Jupiter as the largest stone.

Gamow has a fascinating story to tell, and he tells it well. Though it may be difficult at times for the inexperienced to distinguish between theory and speculation, there is a firm basis of evidence for most of what is said. The book is easy to read, the presentation being clear and lively. Unfortunately, the liveliness is occasionally rather fatuous. As examples, the author must refer to the "smashing results" of some work on nuclear fission, and, in dealing with his own work, started at Göttingen, on radioactive decay, he writes: "Göttingen was a dull little town whose total of entertainment possibilities was represented by two poor movie theatres; the author, who had hoped for something more on his first trip abroad, had nothing better to do than to take up research." Also, why a straightforward Maxwell distribution diagram should be supposed to make a more popular appeal when depicted on a misshapen blackboard on an unstable easel it is difficult to imagine. These, however, are minor blemishes in a book which will be warmly welcomed not only by that wide general public to whom things astronomical make a strong appeal, but also by scientists who have had little opportunity of keeping abreast of the remarkable recent progress in developing still further the relations between stars and atoms. E. C. S.

The Behaviour of Slow Electrons in Gases. By R. H. Healey and J. W. Reed. [Pp. xii+179.] (Sydney: The Wireless Press for Amalgamated Wireless, Australasia, Ltd., 1941. Agents for Great Britain: Messrs. Iliffe and Sons, Ltd., London. Price 20*s.*)

Townsend's application of diffusion methods is one of the high lights in the early history of the investigation of the properties of electrons. These methods have rather fallen into the background, partly, perhaps, owing to the development of experimental methods the results of which are amenable to more straightforward theoretical analysis. Owing to the interest of the diffusion effects themselves, however, the appearance of a book dealing mainly with diffusion methods is very welcome, particularly as the more recent work is accessible only in some~~what scattered original~~ papers.

The experimental methods and the relevant theory are described in the first three chapters. These are followed by a comprehensive survey of the results which have been obtained about the energy of agitation, drift velocity, mean free path, energy losses in collisions, and probabilities of attachment of electrons in different gases. Other methods of investigation are briefly described, and a final chapter, dealing with applications to problems of the ionosphere, includes discussions of the effect of thunderstorms, the production of auroræ, and the interaction of radio waves. There is a full bibliography of the whole field.

Although the experimental results have been carefully sorted out, there is very little discussion of their theoretical significance, and the general reader may find the book rather heavy going. Moreover, for a work of reference, the numerous tables and diagrams are not sufficiently self-explanatory. There is an excessive use in the short captions of symbols whose meaning often cannot be found except by careful search through the text. None the less, for specialists, or for those who wish to make a thorough study of the particular field, this well-produced book will undoubtedly be most useful. E. C. S.

NOTICE TO SUBSCRIBERS TO THE PHILOSOPHICAL MAGAZINE.

Subscribers are advised that for the period of the War, one volume (comprising 12 monthly parts) will be issued instead of two volumes (each comprising 6 monthly parts). The Contents and Index will appear with the December part.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

LX. *The Development of the Trichromatic Theory of Colour Vision.*

By Prof. W. Peddie[*].

[Received April 27, 1942.]

1. The development of the Trichromatic Theory has suffered great retardation through failure to separate clearly between the formal theory itself on the one hand, and a theory of the mechanism in accordance with which the trichromatic activities are themselves manifested, on the other.

The theory in itself makes no assertion regarding the process of colour perception other than the postulate that there are exhibited in it three, and no more than three, independent, *i.e.*, mutually non-interfering, activities.

It makes no assertion, except by way of illustration, regarding the existence or non-existence of three sets of nerve-fibres, or three sets of optically active chemical substances, or any other mode of activity. It acts, in that respect, as a blank formula the details of which are to be settled by further postulation in accordance with observation or experiment. Much good recent work in the latter direction is now accumulating.

In the early development of the theory no assumption was made, except as first approximations, of Newton's linear law of colour mixture, or the so-called Weber-Fechner law of sensation, or Abney's law.

It is desirable that these conditions should be clearly kept in view; for otherwise there may arise, in the course of discussions regarding structural or functional conditions, fruitless affirmations or contradictions regarding the truth of the simple theory of trichromasy itself.

2. The bare statement of the theory has now at last been accepted, with practical unanimity, as a statement of fact—the fact that the matching of colours can be carried out, in normal vision, by means of three suitably chosen, independent, fundamental colours. That being the case, the further conclusion that there cannot be more than three fundamental *sensations* becomes a mere matter of *strict logic*. For, if we adopt the simple forms of Newton's and Fechner's laws, we can write (*cf.* § 18):—

$$\epsilon^{S_1/k_1}=a_1x+b_1y+c_1z,$$
$$\epsilon^{S_2/k_2}=a_2x+b_2y+c_2z,$$
$$\epsilon^{S_3/k_3}=a_3x+b_3y+c_3z,$$

where each triple sum represents the retinal illumination, expressed in terms of our chosen experimental units, which would stimulate one alone of the absolute fundamental sensations S. And if we now assume that a fourth absolute sensation, expressed by

$$\epsilon^{S_4/k_4}=a_4x+b_4y+c_4z,$$

---

[*] Communicated by the Author.

can exist independently, we can eliminate the variables x, y, z amongst the four equations and obtain the relation

$$\mathrm{S}_4=k_4 \log \left(p_1\epsilon^{\mathrm{S}_1/k_1}+p_2\epsilon^{\mathrm{S}_2/k_2}+p_3\epsilon^{\mathrm{S}_3 k_3}\right),$$

where p_1, p_2, p_3 are constants compounded of the twelve given constants a, b, c. But this equation enables us to calculate the additionally postulated sensation in terms of the other three. It is therefore not possible to have more than three independent fundamental sensations, if every possible luminous sensation can be excited by means of three fundamental types of stimulus.

All other forms of colour vision theory, however great may be the number of fundamentals apparently involved (as in those of Hering, Mrs. Ladd Franklin, G. E. Muller, Houstoun, etc.), must be reducible to a trichromatic form by the needed number of relationships existing amongst them. It is a long time since Helmholtz pointed out to Hering that his system is really trichromatic. And it is a long time since he pointed out that two different lights can be made indistinguishable by the variation of either with respect to intensity, whiteness, and hue—three variables only. One form of representation may be more directly adapted to exhibit or explain some particular phenomenon than is another; but, on the whole, the system which uses fewest variables must exhibit greatest simplicity. Hering's system was expressly designed to show the simultaneous vanishing of colour from the two members of a pair of complementary colours when dichromasy existed. But does it really do so more simply than does the fact that, in a colour diagram, two complementary colours lie on opposite sides of the white point?

If psychologists feel compelled to recognize five, or more, psychological sensations, the trichromatic theory is under no need to raise objection. Results which can be expressed in terms of three fundamental stimuli are also expressible in terms of a greater number: but not so readily. More complicated algebraical or numerical work is involved, and more complicated observations have to be made. (See Section 21.)

3. For the development of the theory beyond the stage of mere trichromasy, *but still without reference to structural or operational conditions*, further experiment or observation had to be appealed to. The purpose of the appeal was to find a quantitative law which could lead to the prediction of new results.

Newton's and Abney's Laws.

Newton's law is the fully recognized law for finding the resultant colour and intensity of the light arising from the combination of component lights of given intensities and colours. It is amazing that he could enunciate the law, simple and highly accurate, from the data available to him. For, in consequence of the crudeness of apparatus, he was unable to be quite certain of the existence of spectrum complementaries, or of the composition of white light from three spectrum constituents.

The accuracy of the law is a fine example of the truth that "Facts will always flow in the channel of the deepest speculative perception."

Abney's law is essentially Newton's law in its limitation to three components. If these two laws were strictly accurate, lines of constant colour in Lambert's colour pyramid would be straight lines radiating from the origin. This is actually not the case in general. Thus both laws must be regarded as good first approximations only.

The Weber-Fechner Law.

4. The relation which subsists between stimulation and sensation is of absolutely fundamental importance in the development of any theory of colour perception. Weber's statement of the law is that a small finite change in sensation, due to a small finite change in stimulus, is proportional to the fractional change of stimulus. Fechner's statement extends the law to infinitesimal changes.

The former statement is based on experiment, the latter is a deduction. Exception is therefore sometimes taken to Fechner's form; but, in this particular case, the objection does not apply. For, if Fechner's form be true with a sufficient variety of steps, the other is also true of necessity. It is the integrated expression that is important :—Change of sensation is proportional to change of stimulus. (See further Section 23.)

As in the case of Newton's law, there is a limited range of applicability for this law, and the departure is more violent outside the range than it is in the former case. Newton's law holds for sufficiently small intensities: Fechner's law applies at sufficiently great intensities.

5. In the case of any curve of finite curvature, which corresponds to observation within the limits of experimental error, the tangent to that curve at a chosen origin, taken on the curve, gives a suitable approximate linear relationship, which is observationally *true* to greater or less extent dependent on the observed curvature. And it would be wrong to call it incorrect, or to say that it gave a false law because the curve ultimately departs from it even largely. It gives the first terms of an empirical expansion in a series of powers. Such an expansion, if it involves many terms and cannot be condensed so as to involve few constants, is never regarded by the physicist as the absolute law; for he finds that nature seems to work in accordance with simple procedures.

6. But, in Helmholtz's curve, in which the fractional change of intensity per unit change of sensation was plotted, as ordinate, against the logarithm of the intensity, there was a very considerable range of intensity throughout which the fractional change was closely constant. Ultimately, in consequence of dazzle, as he called it, the fractional change again increased with increase of intensity. He introduced a factor to take account of dazzle, so giving a modification of Fechner's law. This was a quite proper procedure, recognizing that, for *a given step in sensation*, the fractional step in stimulus being greater the factor must be less than unity. He used it in the form $1/(1+px)$ for a stimulus of one type,

where p is small, and x is the intensity—the unmodified form of Fechner's law being employed. Thus

$$dS=\text{constant}=k\left(\frac{dx}{x}-\frac{dx_0}{x_0}\right)=kd\log\left(\frac{x}{x_0}\right),$$

where k is the psychological constant which gives the ratio of the sensation change to the intensity change, and x_0 is the threshold stimulus. This gives

$$x_0=1+px,$$

so that the threshold is variable. (See further Section 22.)

7. In the same connexion Helmholtz used, apart from dazzle, the expression $1/X$, where X is a function of x, as the value of the change in $\log x$ per unit change of sensation. Hence

$$\log\left(\frac{x}{x_0}\right)=\int\frac{dx}{X}.$$

These are two cases in which Helmholtz dealt with a non-constant threshold. In general, in all his work, he postulated that the threshold was constant (which requires X to be proportional to x), a condition usually difficult to attain.

8. In the case of dazzle, with the external stimulus x large, recent work shows that dazzle ceases if x is raised slowly enough to its extreme given value. This at once suggests that dazzle is due to incomplete accommodation, so that the threshold is a function of the time of accommodation, and also, no doubt, of the manner of approach to that condition.

For example, in the case of Helmholtz's simple dazzle factor, the condition

$$p=p_0\epsilon^{-kt},$$

with k constant, would give the required result under exposure for a sufficiently long time to the originally dazzling illumination. The theory is capable of predicting this result, though it can only indicate a possible simple law. Experiment must decide. Here there is a great field lying open for investigation.

The recent experiments alluded to above, the most recent being those of Craik, show that Fechner's law holds, at high enough intensities, not merely throughout the considerable range studied by Helmholtz, but throughout a very large range indeed. The foundation of the psychophysical law on that of Fechner has now a very complete justification. It is a law of conditions in equilibrium, which occurs when accommodation is complete.

The Extension of Fechner's Law.

9. This law is a simple law; and the law of variation from it should be simply expressible also if it is to be of ready practical use. It must be in agreement with observation, and it should be based on a recognizably

acting cause. All of these conditions were carefully attended to in the development of the law by Fechner himself in the first place; and subsequently, in simple but powerful mathematical treatment, by Helmholtz.

Fechner recognized that the eye is normally in a state of excitation even though there is no external illumination falling upon it, and this effect is only made more complicated by external action. He called it the *self-light* of the eye. And he added its amount to the external amount in order to get the total effective illumination. Thus the total illumination was $x+a$, where a is the internally produced part of the excitation, and was assumed to be constant by Fechner. So that the fractional illumination was $dx/(x+a)$, Fechner's unmodified law being still taken as true as regards the *total* illumination.

Helmholtz showed how this first modification of the law led directly to the expression of a deviation from Newton's law of colour mixture, at higher intensities, which was in general agreement with observation.

In further development of the subject he proceeded to take the self-light as variable with the external intensity. Any model of variation could have been postulated, but Helmholtz explicitly postulated that the *observed* peculiarities of the self-light alone must be used.

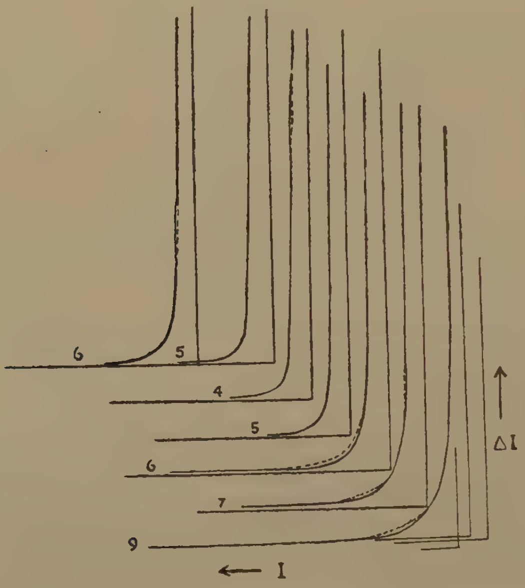
10. Now an observed peculiarity is that the self-light is constantly varying both in respect of its intensity and of the distribution of that intensity over the retina. He expressed this condition in such a way as to obtain the quantity sought in the form of an integral, which, however, could not be evaluated directly as the unknown magnitude itself appeared under the integral sign. He surmounted this difficulty through integration by parts, so as to bring in, with each successive integration, two additional quantities, known to be constant, but not known as to actual magnitude. Yet they are all theoretically determinable by fitment to the data of observation. By sufficient extension of the integration, an adequate number of constants can theoretically be found to compel the fitment to be within the possible errors of observation. The theoretical solution of the problem is thus complete. There is no deficiency in the *theory* in so far as correspondence to *any* observed results is involved.

The total infinitesimal change of sensation due to a change of external illumination on the retina is regarded as the sum of all the changes of sensation caused by that change of external illumination on each element of the retina over which the self-light is closely constant. For each such element Fechner's law is applicable, in its unmodified form, with regard to the mean total illumination.

This extension of his work by Helmholtz makes possible a still fuller account of the observed deviations from Newton's law than Fechner's first modification gave. This also is part of the so-called Young–Helmholtz theory.

The calculation of the pairs of constants, referred to above, becomes very complicated after the second approximation. Calculations carried out by Clara O. T. McKenzie, a former student at University College, Dundee, showed that a very good representation was reached over the whole range of observation with four constants only. A highly reduced copy of one of her diagrams is shown below (fig. 1). Two of these curves are in duplicate and exhibit the effect of a change in the value of the parameters. The applicability of the formula is evidently very high. It shows that the simple unifying self-light postulate takes account well of enormous variation from the condition of equilibrium specified by the unmodified psychophysical law as given by Weber or Fechner. And the test is even more searching than appears; for the calculation of the constants was simplified by taking a pair of the fixing points well away from the region of great curvature on both sides.

Fig. 1.



Ordinates in the diagram represent the smallest fractional change of intensity which is just perceptible, while the abscissæ give the intensity itself. Dotted parts follow observational results.

The Colour Vector Development.

The influence of the development of Fechner's law upon that of Newton's law affects also Abney's law, which can only be regarded as a good first approximation. The new point which arose at this stage of Helmholtz's work was the question of the law of combination of the three fundamental sensations. Is this combination law independent of Fechner's law and of Newton's law, which the generalized Fechner's Law itself modifies? Helmholtz had strongly expressed doubts as to the validity of estimates of equality of brightness in differently coloured lights. He felt that these estimates were not independent of the qualitative difference in hue.

12. In respect of phenomena which are expressible in terms of three independent variables, not more than two *à priori* laws can be laid down. When these are laid down, the theory predicts a third law.

Now Helmholtz's outstandingly great contribution to the trichromatic theory of colour vision was the idea of colour vectors in a three-dimensional space. And he had specially developed Fechner's law himself. It was therefore natural for him to select Fechner's modified law, together with a vector additive law, as the two laws to be explicitly represented. In doing that he naturally, and from his own point of view, rightly, as being simplest, chose the Euclidean law for his colour sensation vector space.

The Differential Sensitivity Test and the Absolute Fundamentals.

13. Helmholtz gave three tests by which the suitability of his vector law could be tested. He carried out one, and was satisfied with the result; and he indicated another, which does not seem to have been carried out, although he referred to it as a stronger one. A third was carried out for him, and the result also satisfied him at first. He worked out an expression for the differential sensitivity throughout the spectrum, based upon Fechner's unmodified law and his own sensation vector law. And he pointed out clearly that the differential sensitivity law, being non-linear in the three independent variables, could be used for the determination of the fundamental variables in the absolute sense. This determination was made for him; but too great stress must not be laid on the result, for the data used were not, as he indicated, quite free from deficiency. But the general form of the curve thus found, and its location in the spectrum, agreed fairly well with observation. Undoubtedly the general form of the curve depends on the form of the so-called "sensation curves." But that in no way detracts from the success of the theory in predicting an experimental result, for it has no *à priori* knowledge of the form of these curves. The theory fits. But it does not follow that the employment of a wrongly postulated vector combination law may not give rise to obvious disagreement. It may lead to the condition that the determination of the absolute fundamentals, even with heavy labour, may be no easy matter. This determination might scarcely be worth attempting except by way of mean sensation curves obtained from a large number of non-anomalous trichromats. It must be remembered that, before any conclusion regarding this test can be arrived at, linear transformation to the absolute fundamentals must be carried out, so that a Euclidean expression for the magnitude of the resultant can be found.

Here also lies open a great region for experimental and computational investigation. It is a Helmholtzian region, even if it be carried out by means of the unmodified Fechner's and Abney's laws, leaving open for determination by the theory the corresponding expressions for the component *sensations.*

The Pauli-Schrödinger Development.

14. It was first pointed out by W. Pauli that the vector stimulus space which should be used, to give effect to the employment of Abney's law, is that in which the square roots of the stimuli, instead of the stimuli themselves, appear as co-ordinate quantities. In this way, using Abney's and Fechner's laws, the corresponding expressions for intensity, hue, and saturation can be evaluated. This was carried out by Schrödinger in Cartesian notation.

A quaternionic investigation is given in 'Nature,' cxxiv. p. 791. Intensity is given by the square of the radius vector: saturation by the angular distance, θ, of the colour vector from the white vector: and hue by the product of θ into ψ, which is the angular distance between the plane of the white vector and the colour vector on the one hand, and the plane containing the white vector and the dominant fundamental vector on the other. Both angles are to be measured in terms of $\pi/2$ as the unit.

Schrödinger omitted the factor θ in the case of hue by regarding colour sensation as vanishing in the close neighbourhood of the white vector in consequence of the colour stimulus falling below the threshold there.

Difficulties Felt Regarding the Theory.

15. These seem to be due in general to unnecessary limitation of the scope of the theory.

The most fundamental are those in which the validity of a fundamental law is brought into question. In all cases these laws are approximations. In the preceding paragraphs, therefore, attention has been specially directed to the wide extent of the range throughout which each approximation holds, and the conditions under which it applies. If discrepancies are found between observation and approximate results, attention should first be turned to the higher approximations.

Other difficulties arise because of the presence of internal variables. The theory asserts merely that the independent external variables are not more than three in number. These three may occur internally also, along with a multitude of others. For example, the time which has elapsed since the external stimulus began, the nature of precedent stimulation and its magnitude, the interval since food was taken, the amount of precedent exertion, physical or mental strain of observation, etc. Much work has been carried out in recent years in connexion with such points: but the results are seldom given, in a common systematization, by arrangement in the Helmholtzian receptacle log (i/a), in which a is the threshold value of the stimulus and is a function of the various "internal" conditions.

Another type of difficulty relates to the cross connexions which arise amongst the normally independent trichromatic paths or centres in the cases of monochromasy, dichromasy, and many instances of the so-called anomalous trichromasy. An interesting case of the latter type stands apart, and will be discussed later. (Section 21.)

16. Here it must not be overlooked that there is no *à priori* necessity that lapse of action of one of the centres should correspond to dichromasy. The linearity of Newton's law enables us always to specify a set of interconnexions which will originate lapse-dichromasy. Observation of some non-linear effect, as Helmholtz insisted, alone can decide if these centres are the absolute centres. But here there exists a great gap, requiring hard and skilled labour before it can be filled.

Helmholtz showed how to express and use the occurrence of a failing colour which need not be an absolute fundamental colour.

17. A special difficulty has been expressed in connexion with the fact that, with certain fundamentals, the blue representative centre is not influenced by red light of longer wave-length than about 650 μm., where the green sensation ceases as well as the red. How then can yellow light seem to be more saturated after exposure of the dichromatic eye to blue light? Or how can dichromatic yellow-blue vision give ready after image action?

This is a case of very great interest, as also are the problems set by it. If sensation depended only on external stimuli the theory would not be sufficient, provided that the fundamentals were known to be absolute. Helmholtz did not himself investigate any such problem. But (Section 7) he expressed the existence of the threshold though he worked with problems in which it was taken as constant. Had he been asked this question there can be little doubt that he would at once have ascribed the peculiarities to threshold conditions. The sensation curves are fixed only if the thresholds are fixed.

The thresholds of red and blue light are functions, in general different, of the two wave-lengths. The theory is at no loss, but experiment is required to furnish the data.

A still more striking case is one in which a single range of colour alone was recognized in the spectrum. This extended from the normally red end of the spectrum to the green region. Strong yellows, greens, and blues were matched with greys. The violet end of the spectrum was called colourless also; but it was hesitantly said to be, "if anything, red," when attention was specially directed to that end.

But this was not a case of monochromasy. It was a true case of dichromasy with a failing colour near the yellow-green part of the spectrum, and the complementary failing colour near the remote violet region. This was directly made evident by the fact, very easily demonstrated, that, while light from the red region of the spectrum gave no visible after image, the seemingly colourless green light gave a strong "red" after image. The case was one of high-threshold, yellow-blue dichromatic vision.

All these problems, with similar peculiarities, seem to come readily within the sweep of Helmholtz's developments of the trichromatic theory.

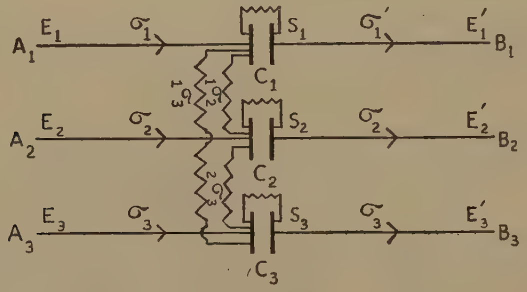
18. A very illustrative electrical model of the trichromatic aspect of the activities which appear in the phenomena of colour vision can be readily constructed (fig. 2).

Take three conducting circuits $A_1C_1B_1$, $A_2C_2B_2$, $A_3C_3B_3$, each containing respectively a condenser C_1, C_2, C_3. Let these condensers have their own plates joined by leakage circuits s_1, s_2, s_3, while their left-hand plates are cross-connected by conductances ${}_1\sigma_2$, ${}_2\sigma_3$, ${}_1\sigma_3$.

The suffixes of any one of these are, of course, interchangeable ; so that ${}_1\sigma_2={}_2\sigma_1$, etc.

If the conductances to the left of the condensers are σ_1, σ_2, σ_3, while those to the right are σ_1', σ_2', σ_3', and if electromotive forces E_1. E_2, E_3,

Fig. 2.



and $E_1,'$ E_2', E_3' act in these respectively, we, without loss of generality, can put

$$E_1'.\sigma_1'=E_1(\sigma_1-{}_1\sigma_2-{}_1\sigma_3)+E_2.{}_1\sigma_2+E_3.{}_1\sigma_3,$$

$$E_2'.\sigma_2'=E_1.{}_1\sigma_2+E_2(\sigma_2-{}_2\sigma_3-{}_2\sigma_1)+E_3.{}_2\sigma_3,$$

$$E_3'.\sigma_3'=E_1.{}_3\sigma_1+E_2.{}_3\sigma_2+E_3(\sigma_3-{}_3\sigma_1-{}_3\sigma_2).$$

The function of the leaky condensers is to enable finite stationary flow, under given stimuli, to occur. The quantities E represent retinal stimuli ; the quantities E′ correspond to cortical stimuli.

Since we are only concerned with ratios of the electromotive forces, no generality is lost by putting

$$\sigma_1'=\sigma_2'=\sigma_3'=1.$$

The six quantities σ_1, ${}_1\sigma_2$, etc., then give the nine coefficients of transformation from the arbitrary fundamentals E_1, E_2, E_3, to the equivalent expressions in terms of the absolute fundamentals E_1', E_2', E_3', or conversely.

Example 1.—Put $E_2=0={}_1E_3$. Then the absolute components are given by

$$E_1'=E_1(\sigma_1-{}_1\sigma_2-{}_1\sigma_3),\ E_2'=E_1\,{}_1\sigma_2,\ E_3'=E_1\,{}_1\sigma_2.$$

It is important to note that E_1' may be either positive or negative relatively to E, in accordance with the sign and magnitude of the bracketed quantity.

The points A do not necessarily correspond to points located on the retina.

Example 2.—Let there be only two external stimulations, E_1 and E_2. The equations are

$$E_1' = E_1(\sigma_1 - {}_1\sigma_2 - {}_1\sigma_3) + E_2 \cdot {}_1\sigma_2,$$
$$E_2' = E_1 \cdot {}_1\sigma_2 + E_2(\sigma_2 - {}_2\sigma_3 - {}_2\sigma_1),$$
$$E_3' = E_1 \cdot {}_1\sigma_3 + E_2 \cdot {}_2\sigma_3.$$

These give

$${}_1\sigma_2 \cdot E_1' - (\sigma_1 - {}_1\sigma_2 - {}_1\sigma_3)E_2' = E_2[{}_1\sigma_2{}^2 - (\sigma_1 - {}_1\sigma_2 - {}_1\sigma_3)(\sigma_2 - {}_2\sigma_3 - {}_2\sigma_1)].$$
$${}_1\sigma_2 \cdot E_2' - (\sigma_2 - {}_2\sigma_3 - {}_2\sigma_1)E_1' = -E_1[{}_1\sigma_2{}^2 - (\sigma_1 - {}_1\sigma_2 - {}_1\sigma_3)(\sigma_2 - {}_2\sigma_3 - {}_2\sigma_1)].$$

Hence
$$E_2' = E_1' \frac{\sigma_2 - {}_2\sigma_3}{\sigma_1 - {}_1\sigma_3},$$

so that the two cortical stimuli will have opposite signs if $\sigma_2 > {}_2\sigma_3$, $\sigma_1 < {}_1\sigma_3$, or conversely.

Both cortical stimuli will be negative if $\sigma_1 \ngtr {}_1\sigma_3$ and ${}_2\sigma_3$ is sufficiently large, provided that we choose E_1 as the largest of three retinal stimuli. The third cortical stimulus is positive.

These two examples show how, in cross connected electric circuits, terminal stimuli may be either like or opposite in sign to the initial stimuli. This illustrates a point which is often felt to form a difficulty.

19. The general case in the working of the model is represented by the equations

$$E_1' = E_1(\sigma_1 - {}_1\sigma_2 - {}_1\sigma_3) + E_2 \cdot {}_1\sigma_2 + E_3 \cdot {}_1\sigma_3 + e_1',$$
$$E_2' = E_1 \cdot {}_1\sigma_2 + E_2(\sigma_2 - {}_2\sigma_3 - {}_2\sigma_1) + E_3 \cdot {}_2\sigma_3 + e_2',$$
$$E_3' = E_1 \cdot {}_1\sigma_3 + E_2 \cdot {}_2\sigma_3 + E_3(\sigma_3 - {}_3\sigma_1 - {}_3\sigma_2) + e_3',$$

where each small e' indicates an internally produced cortical stimulus from which arises the impression of "self-light." Suppose that we arrange $E_1 > E_2 > E_3$. It is evident that the strongest adverse effect, in any one E circuit, with given σ's and negative bracket, is associated with high magnitude of that E, as is to be expected.

And the self-light effect has more influence, in a given circuit, the greater is the total cross conductance from that circuit. That is to say, *A rise of Threshold by the action of Self-Light in a circuit is less pronounced the more complete is the insulation of that circuit.*

This is in accordance with the view that the development of the eye as a light perceptive organ was from complete fusion of paths towards complete insulation of paths. In the latter case, the fundamentals would be absolute in the sense in which Helmholtz used the word.

20. The actual procedure in the optico-cerebral circuits is supposed to arise from liberation of electrons therein in consequence of a transformation of incident light energy into electrical energy by photo-electric action. Under maintenance of the incident illumination, flow of the electrons from the retinal end to the cortical end of the circuits takes place. In the above-illustrated model, the leaky condenser may represent the synapse in the actual circuit. At the cortical end a new transformation of energy takes place, the electrical energy being changed into the form of energy issuing in cerebral activity, as evidenced in the visual sensations.

The law of stability in any pair of directly related interactions is expressed by the statement that if an action P gives rise directly to an action Q, the action Q itself opposes the action P which gives rise to it. In the present case the incidence of light gives rise to the evolution of electrons at the retinal end of the optical fibres. If independent production of electrons takes place, the direct luminous effect will be opposed thereby. In other words, increased intensity of the incident light will be required to give the same cortical response. Or we may say that the internally originated development of electrons necessitates less externally originated development in order to produce the same sensation. The brain cannot distinguish between the origins. So, if the externally produced sensation of luminosity be ascribed to "light," it is entirely consistent to ascribe the internally produced sensation to "self-light."

The term *self-light* has no other implication in the development of colour vision theory than that it indicates the existence of an internally produced action of the same nature as the action to which external illumination gives rise. It is quite independent of the question of trichromasy and its sufficiency in colour sensation. The postulate of photo-electric action lends itself readily to the idea of an internal activity of the needed type.

The existence of the "self-light" effect necessitates an increase of the "threshold value." There is a definite relation between the two which depends on the form of the psychophysical law.

The spontaneous development of electrons may not be so great as to give rise to over-threshold stimulation, and therefore to the sensation of luminosity, but its deterrent influence on perception is nevertheless effective.

On the other hand, I do not find it possible to take the view that it cannot be of sufficient strength to account by itself for all observed phenomena of the type involved in the discussion. After-images can be so bright that they can be strongly visible against a bright sunlit background. By looking first through almost closed fingers, and then gradually opening them, I found finally that I could gaze directly and steadily at the mid-day sun in a clear summer sky. The induced feeling of weakness of perception was very intense. The disk of the sun seemed very dull and of a bluish tinge. That was probably due to contrast with the intensely bright yellowish light of the sun's rim when the involuntary slight motions of the eye exposed an unfatigued part of the retina to it. On then looking at the landscape not the faintest tinge of colour was visible. The blue sky was as grey as ashes, the blue sea seemed a darker grey, green grass or yellow fields all seemed to be grey of a lighter or darker shade. After a few minutes the natural colours became visible again ; first blue, then green, etc.

But the after images of the sun remained visible, even against a brightly sunlit roadway, and were strongly evident during three weeks. Self-light can be intensely strong.

Fechner worked with intensely bright artificial light when he made his pioneering investigations. Newton investigated the action of direct sunlight so effectively that the after images did not disappear for six months. And during the latter part of that time, he had carefully to refrain from even thinking about the matter ; for, if he did so, the images appeared at once. This is an extremely interesting example of reaction with regard to the internal energy which is directly involved in thought. Annulment of the electronic charge at the cortex gives direct rise to the development of thought : therefore independent thought production causes increase of the charge, which is equivalent to increase f self-light.

The Difficulty regarding Perception of Yellow.

21. A very notable fact of direct observation in the case of the great majority of observers is the characteristic quality of the sensation produced by the incidence of yellow light on the retina. The impression seems to be as unique as either of the impressions produced by red or green light.

But there is no fact more certainly known than that of the fallibility of many direct deductions from the evidence of the sensations. In each case there is a physical explanation of the discrepancy, *e. g.*, in the case of the ordinary mirage. Yet the explanation seems to lie at hand in the fact that some observers, whose vision is fully trichromatic, actually perceive red and green in yellow light, and can as readily speak of red-green light as they can speak of blue-green light. These observers have hitherto been classed as abnormal trichromats. They are really the most highly developed trichromats.

According to the theory of the gradual development of trichromasy in the slow progress of time, that development proceeded from monochromasy, by way of dichromasy and imperfect trichromasy, towards the final stage. The isolation may have proceeded in a variety of ways, so that there might be different types of monochromasy and dichromasy. There might even have been no completely dichromatic stage. The question is this. How can we explain the existence of two types of perception, in one of which, *e. g.*, blue-green, a compound sensation is experienced, while in the other the sensation is single, *e. g.*, yellow ?

The electrical model already considered may give an indication of the solution. If a direct short circuit occurs at two of the cerebral termini, one and only one compound sensation will be experienced. If the isolation becomes more developed, a variation in the sensation, as becoming greener or bluer in accordance with the wave-length of the stimulating light, will be observed. An electro-optical model could be constructed which would actually exhibit the difference between the two modes of action. If this could be regarded as supplying a true analogy, we might conclude that eyes sensitive to the red-green impression in the yellow region of the spectrum, and to the blue-green impression in the blue region, are more highly developed than the average eye. This may be true in at least some of the cases of "red-green" discrimination in spectral yellow.

Numerator versus Denominator.

22. In experiments which deal with the least change of sensation which is just discriminable, we have absolutely no means of judging if two such separate steps are really equal. Because we have no choice in the matter, we define them as being equal. The case is very similar to that which arises in thermometry when we define equal intervals on a linear scale as corresponding to equal intervals of temperature. In the case of thermometry we can test the accuracy of the definition by experiment. In the case of colour vision no experimental test is possible. We have no knowledge of the loss of energy in the final transformations which are involved.

The value of the threshold, as we have seen, depends upon many conditions, and a great amount of valuable work has in recent years been carried out in connexion with some of these conditions. It is therefore a matter of great importance that uniformity should be aimed at in the formal treatment of the results of observation, so that ready comparison of the results of different observers and theorizers may be made.

It is of great importance in this respect that internal causes of variation of the threshold should be clearly separated from the external cause. And this is specially important in those cases in which external conditions have influence on internal effects, or when two different external actions each influence the effects of the other.

A very interesting example of the latter type (referred to by Dr. W. D. Wright in his paper published in the Proc. Roy. Soc, B, cxv. p. 84) has been given by Schjelderup, who found that the minimal change of threshold due to a central illumination of intensity i along with a surrounding illumination of intensity I was given by the expression

$$ai+b\mathrm{I}.$$

In so far as i is concerned this is a modification of Weber's formula. But it is really the proper extension of Weber's formula to include the case of multiple external illumination.

If our object were to express the effect of i, *regarded as producer of influence on the central retinal image*, we should express the effect of the surrounding illumination as an effect upon the threshold of i. The appropriate Weber–Fechner (see next section) formula is

$$dS = kd \log (i/m),$$

where m is the magnitude of the threshold. So, using Schjelderup's experimentally found value for this change of threshold due to the establishment of the illuminations I and i, and their maintenance until equilibrium is attained, we find

$$\log m = (\log m_0 - ai) + (1 - bI) \log i,$$

where m_0 is the value of the threshold immediately before the initiation simultaneously of the two illuminations.

What is Fechner's Law?

23. Undoubtedly it is true that Weber's law expresses very correctly, over a wide range, the constancy of the finite but small fractional change of sensation which accompanies a small just observable step in stimulus. But, as already noticed (Section 4), if Weber's form is true, so also is Fechner's differential form. This cannot be too strongly stated. For, if a set of Weberianly related points A, B, C, are observed along the curve of fractional stimulus plotted against time, in company with an exactly similar set A′, B′, C′, ... displaced only infinitesimally from the former, the characteristic property of the curve necessarily gives the same fractional ratio for the infinitesimal steps as for the wider steps.

Consequently, it is as correct to employ Fechner's differential form of the relation as Weber's small-step form; and therefore also just as correct to use the integrated form with its essential introduction of the threshold value of the stimulus—so enabling us to distinguish completely between the effects of the change of an external stimulus, *regarded as an independent agent*, and its effects *in so far as it influences the threshold directly*.

And it must be carefully kept in view that the law only applies in the condition of equilibrium when accommodation is fully attained. The electrical model illustrates this well. After the steady condition of kinetic equilibrium is attained, any change of an E may give rise to changes of the condenser potentials, or to inductive action which causes reverse electromotive forces. But the theory is adequate to the investigation (through experiment) of the laws of variation in all such cases.

In particular the further possible development of the theory invites its use in the solution of many present problems—such as the question of a limit of intensity beyond which the simple form of the law has to be altered whether in consequence of dazzle or in consequence of the entry of new physical or chemical or physiological conditions.

As an illustration, the accompanying set of curves (fig. 3) may be compared with observational curves obtained, in the progress of adaptation, by Dr. W. D. Wright, especially those exhibited in fig. 11 of his paper on the Foveal Light Adaptation Process (Proc. Roy. Soc., B, cxxii. 1937).

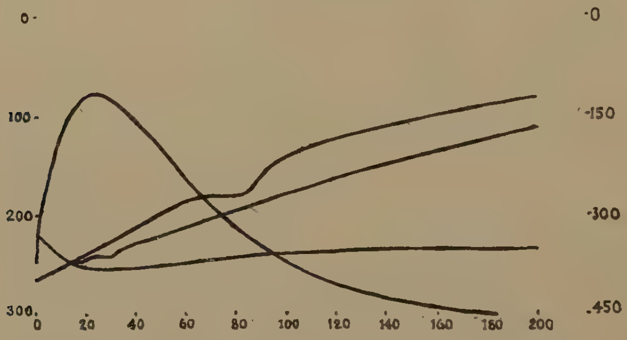
The ordinates, y, are measured downwards from the zero level while the abscissæ give time in seconds—as in Dr. Wright's diagrams. The left-hand scale refers to the pair of curves which exhibit double contrary flexure. But no one of the four curves is intended to do more than exhibit trends similar to those observed. I have no doubt that fair correspondence could be obtained by suitable choice of the constants. And the number of constants is only four in the case of the curves of single contrary flexure, and only five in the case of the other two.

The equation employed is of the form

$$y = A + (a + bt + ct^2) . e^{-pt},$$

with $c=0$ in the two simpler cases. The quantity A merely represents the position of the asymptote when accommodation is complete. The time function in brackets gives the instantaneous value of the effect of the external stimulus, which at once begins to decay in obedience to the characteristic law of exponential variation with time. It is of interest, and very suggestive, that two curves, so different as those in which $c=0$, can exhibit a stimulative influence which is linear in time.

Fig. 3.



Dr. Wright's fig. 5, in the paper referred to above, gives direct testimony to the accuracy of Fechner's law as expressed in the Young–Helmholtz theory. For it gives

$$\log I - \log I_0 = A + B \log I_0,$$

where A and B are constants. But Fechner's law gives, in its fullest differential form, the condition

$$dS = k(dI/I - dI_0/I_0),$$

which reduces to the simplest form

$$dS = k(dI/I)$$

when I_0 is constant. But I_0 constant is the condition that accommodation is complete.

If the bracketed quantity in the second last equation is approximately zero, it might be entirely wrong to conclude that sensation is independent of change of I.

Thus Nutting's data furnish a *proof* that, as a good first approximation, in the equilibrium condition, attained after complete accommodation has been reached (this condition being attested through the constancy then reached by the threshold value of the stimulus), *Fechner's law is verified.* No violation of the law is shown by a departure from it under incomplete accommodation, provided that the well-known limits of too low or too high intensity are not exceeded.

Conclusion and Appeal.

24. At last, within recent years, nearly a century and a half after its first promulgation, the Young–Helmholtz trichromatic theory stands established by a wealth of experimental evidence. Much development remains to be made. And I trust that full use of its great power as an aid in research will meet with increasingly full reward.

LXI. *The Design of Pole Pieces of Electromagnets and the Forces acting on Small Bodies placed in their Magnetic Fields.*

By N. Davy, University College, Nottingham *.

[Received April 20, 1942.]

Abstract.

Air gaps between electromagnet pole pieces of various shapes are considered from the point of view of the mechanical force on a small para- or diamagnetic body placed at any point in them. Assuming the face of a pole piece to be a magnetic equipotential surface, the theory of conformal transformations is applied to the two-dimensional figure formed by a transverse section of the pole pieces and air gap. Faces whose sections are given by the curves $r^{\frac{3}{2}} \cos\frac{3\theta}{2} = \text{constant}$ are shown to produce a field such that the mechanical force on a small weakly magnetic body is constant in magnitude everywhere in the air gap. The direction of this force is along a radius vector drawn from the origin of co-ordinates to the point in question. Such a field is of use in susceptibility measurements with a torsion balance. Other cases considered are (*a*) a convex circular cylinder facing a plane pole piece, (*b*) a concave circular cylinder facing a plane pole piece. Some experimental tests of the theory are described. The force on a small weakly magnetic body placed near a short bar-magnet is calculated, and curves of motion of such a body, analogous to, but distinct from, lines of force are obtained.

---

* Communicated by Professor L. F. Bates, Ph.D., D.Sc.

Introduction.

SINCE the magnetic permeability of a ferromagnetic is large compared with unity, the surfaces of the pole pieces of electromagnets can be regarded as equipotential surfaces. This view is supported by such authorities as J. J. Thomson (1), and has been used by Fereday (2) in formulating a theory of the shape of pole pieces of electromagnets giving a desired type of magnetic field in the air gap between them. If this assumption be sound, it would seem to bring the theory of the magnetic fields between pole pieces within the scope of those mathematical methods which have already proved so successful in electrostatics, current electricity, heat and hydrodynamics. The particular method applied here is that of conformal transformations, suitable for two-dimensional systems.

Pole Pieces with a Field giving a Constant Mechanical Force.

In measuring the magnetic susceptibilities of para- and diamagnetic substances by torsion balance methods it is usually assumed that a specimen, placed in a magnetic field, experiences a mechanical force of value $\frac{(k-k_a)}{2}\frac{\partial H^2}{\partial s}$ dynes per c.c. in any direction s, where H is the intensity of the field, k and k_a are the volume susceptibilities of the body and surrounding medium, respectively. One of the chief difficulties to be overcome is that $\frac{\partial H^2}{\partial s}$ alters as soon as the body moves away from its first position. Curie and others have placed the body at a point of maximum value of $\frac{\partial H^2}{\partial s}$. Fereday (*loc. cit.*) used spherical pole tips to make $\frac{\partial H^2}{\partial s}$ constant over a distance of 1·5 cm. in a certain direction. His pole piece system has circular symmetry about an axis normal to the faces of the poles. Bates, Baker and Meakin (3) tried to avoid this limitation, which would not permit the examination of the magnetic properties of amalgams, which separated into layers of different density, by the use of cylindrical pole pieces, and the present work arose from a discussion as to whether the particular pole pieces used by them could be replaced advantageously by others of different curvature. It is now proposed to attempt to find a two-dimensional system of pole pieces, such that the force on a small weakly magnetic body, placed in their air gap, is constant at all points. To apply the method of conformal transformations we have to guess an appropriate transformation. In doing so, we take a hint from the condition we are trying to establish. Since $\frac{\partial H^2}{\partial s}$ is to be a constant, write $\frac{\partial H^2}{\partial s}=a$, cross multiply by ∂s and integrate. This gives $H^2=as+b$, where b is a constant of integration, which, we assume, can be arranged to be zero. Then if $H^2=as$, $H=\pm a^{\frac{1}{2}}s^{\frac{1}{2}}$. H is not generally in the direction s, but supposing it were, we could put $H=-\dfrac{\partial U}{\partial s}=\pm a^{\frac{1}{2}}s^{\frac{1}{2}}$, where U is the potential. Integrate again and we get $U=\pm\dfrac{2a^{\frac{1}{2}}s^{\frac{3}{2}}}{3}+c$, where c is a constant. In such a case the variable part of U would be proportional to $s^{\frac{3}{2}}$. This rough calculation suggests that we ought to try a transformation involving the index $\dfrac{3}{2}$. Accordingly we try the transformation $W=Cz^{\frac{3}{2}}$, *i. e.* $U+iV=C(re^{i\theta})^{\frac{3}{2}}$, where U can be regarded as the magnetic potential, V the conjugate function, C a constant, r and θ the polar coordinates of a point in the z-plane. Hence

$$U+iV=Cr^{\frac{3}{2}}e^{\frac{i3\theta}{2}}$$

$$=Cr^{\frac{3}{2}}\cos\frac{3\theta}{2}+iCr^{\frac{3}{2}}\sin\frac{3\theta}{2}$$

and $$U=Cr^{\frac{3}{2}}\cos\frac{3\theta}{2} \quad\text{and}\quad V=Cr^{\frac{3}{2}}\sin\frac{3\theta}{2}.$$

Taking U as the potential, the equipotential curves are given by $r^{\frac{3}{2}}\cos\dfrac{3\theta}{2}=\text{constant}$, and the lines of force by $r^{\frac{3}{2}}\sin\dfrac{3\theta}{2}=\text{constant}$. Some of these curves are represented in fig. 1 and figs. 4, A 1, A 2, B 1, C 1 and D 1. Any pair of the equipotentials will serve as opposite pole faces. The lines of force have the same shape as the equipotentials but are rotated through 60° about an axis perpendicular to the plane of the diagram.

The next step is to calculate $\dfrac{\partial H^2}{\partial s}$ at any point and also to find its direction. This is done by calculating first its components along and perpendicular to any radius vector OP. Call these components F_r and F_θ. Let the magnetic field at any point P be H and let its components along and normal to OP be H_r and H_θ, respectively.

Then $$H^2=H_r^2+H_\theta^2.$$

Also $$H_r=-\frac{\partial U}{\partial r}=-\frac{3}{2}Cr^{\frac{1}{2}}\cos\frac{3\theta}{2}$$

and $$H_\theta=-\frac{\partial U}{r\partial\theta}=\frac{3}{2}Cr^{\frac{1}{2}}\sin\frac{3\theta}{2}.$$

Hence $$H^2=\frac{9C^2r}{4}.$$

Next $$F_r=\frac{\partial H^2}{\partial r}=\frac{9C^2}{4} \quad\text{and}\quad F_\theta=\frac{\partial H^2}{r\partial\theta}=0.$$

Hence $$\frac{\partial H^2}{\partial s}=\{F_r^2+F_\theta^2\}^{\frac{1}{2}}=F_r=\frac{9C^2}{4}.$$

The resultant force per c.c. $=\dfrac{\partial H^2}{2\partial s}(k-k_a)=\dfrac{9C^2(k-k_a)}{8}$ dynes per c.c. if e.m.u. are used for C, k, and k_a. Hence, if k and k_a are independent of

H, the resultant mechanical force on a small weakly magnetic body at any point P (fig. 1) is constant in magnitude and is directed along the radius vector OP. It is directed away from or towards the origin according as k is greater or less than k_a.

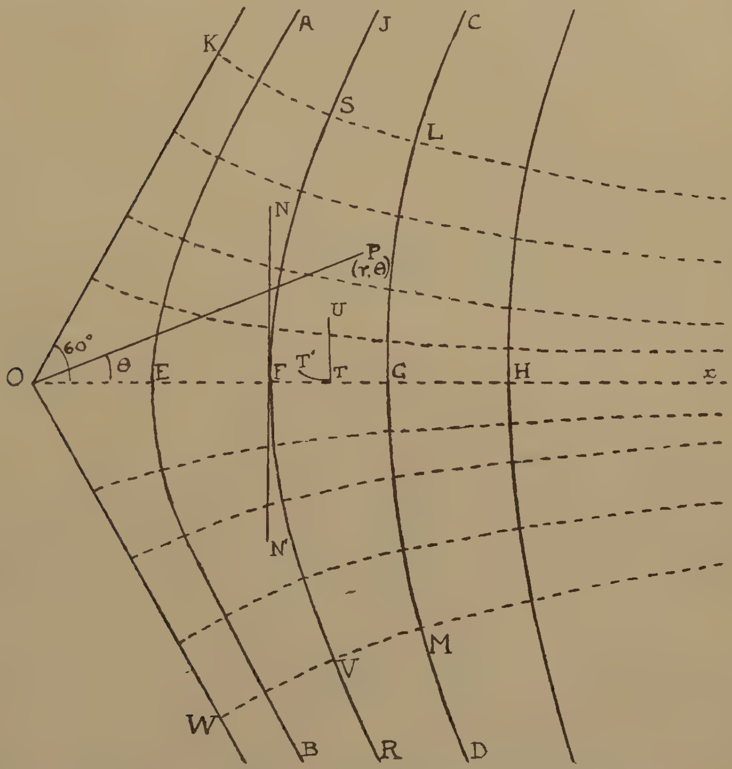
The problem just solved is a two-dimensional one. The corresponding three-dimensional system would have pole pieces in the form of infinite cylinders, with vertical generators normal to the plane of fig. 1. Thus figs. 1 and 4, A 1, A 2, B 1, C 1 and D 1 would represent horizontal sections of some of these cylinders, and the curves should extend to infinity in the plane of the figure. In a real electromagnet the pole pieces would be finite in depth and also in cross-section. To approximate to the above, the pole pieces could be made as deep and wide as possible, and the body placed centrally and half-way down. The shape of the horizontal sections of the faces of the pole pieces could be made to fit any pair of equipotential curves, *e. g.* curves AEB and CGD of fig. 1. To the left of AEB and to the right of CGD would be the metal of the pole pieces, with the air gap between. The field of force shown in fig. 1 bears a considerable resemblance to that selected by Fereday (2) in his first paper, p. 252, as likely to give a constant mechanical force on a small body.

If in a torsion balance experiment the small specimen was suspended at a point F on the geometric axis OG of pole pieces KOW and LGM (fig. 1), edge effects would be small. The axis of rotation of the balance could be arranged to intersect the horizontal plane of the figure at some point such as N or N′, along NN′, a normal to OG through F. Then, since OG is a tangent to any circle whose centre lies on NN′, the body will not move far from Ox during a small angular rotation about N or N′. The fractional decrease in the moment of the mechanical force on a body originally at F, when it rotates through a small angle α radians about any point on NN′, is $\frac{\alpha^2}{2}$, which is about $\frac{1}{6565}$ per square degree or ·0152 per cent.

Experimental Tests.

By kind permission of Professor L. F. Bates, and the friendly zeal of Mr. E. M. Somekh, one of his research students, the above theory was tested as follows. Mr. Somekh filed out magnet steel pole pieces with profile curves, in horizontal section, of the form $r^{\frac{3}{2}} \cos\frac{3\theta}{2} = \text{constant}$. In appearance the sections of their faces bore a close resemblance to curves RFJ and DGC of fig. 1, or the curves in fig. 4, D 1. These were mounted at the proper distance apart on a large Rutherford electromagnet. The magnetic intensity H at various points in the air gap between F and G was measured with a fluxmeter. The results were found to be in accordance with the law that H^2 is proportional to r, r being measured from O. They were not in accordance with a number of other plausible relations connecting H and r that were tested. Next, the mechanical torque on a mass of pure mercury suspended initially near a point T on the line FG was measured by a delicate torsion balance method. The axis of rotation of the balance intersected the plane of fig. 1 at a point U along a normal to FG through T, and the torque was measured when the mercury was at various points along a circular arc TT′ touching FG at T. The calculated torque should only fall off by about 1·52 per cent. during a rotation of 10 degrees about U. The experimental measurements, which were liable to errors not exceeding ±5 per cent. indicated that the torque remained constant over this range.

Fig. 1.



Equipotentials or Pole Pieces ———, Lines of Force - - - - - .

As an extra test the torque on the mercury was calculated, using the constant $\frac{9C^2}{4}$ measured by the fluxmeter experiment, assuming $H^2 = \frac{9C^2 r}{4}$ along FG, and measuring r from O. When the mercury was at T the torque on the system was $\frac{(k-k_a)}{2} \cdot \frac{\partial H^2}{\partial r} \cdot v$ multiplied by the arm UT, where v=volume of the mercury. Writing $\frac{\partial H^2}{\partial r} = \frac{9C^2}{4}$ the torque was $\frac{9(k-k_a)C^2 . v . \mathrm{UT}}{8}$. This came to 1·41 in certain arbitrary units. This value was compared with that obtained by an electrodynamometer method. The moving coils of the electrodynamometer were carried on the torsion balance and the same adjustable electric current could be sent through them and a pair of neighbouring fixed coils so as to apply a calculable torque to the balance system. In this way a torque equal to the magnetic torque discussed above could be produced, equality being assumed to exist when the mercury remained at T. The value of the torque calculated from the dimensions of the coils of the electrodynamometer, which could only be done roughly, came to 1·59 in the same arbitrary units as before. The qualitative agreement was regarded as further support for the above theory. The elementary method of sprinkling iron filings on a card placed over the pole pieces also revealed lines of force of the kind expected.

Further Comments on this Case.

(*a*) Should the specimen contain a fragment of a ferromagnetic impurity, for example, in a liquid amalgam, the above theory seems to indicate that the ferromagnetic will move along a radius vector OP from the origin O to that side of the container remote from the origin, but the force and the moment of the force about the axis of the torsion balance will not be altered thereby.

(*b*) Since $H^2=\frac{9C^2r}{4}$, $H=\frac{3Cr^{\frac{1}{2}}}{2}$ and $\frac{\partial H}{\partial r}=\frac{3Cr^{-\frac{1}{2}}}{4}$, and lines of constant H are circles $r=$constant.

(*c*) The magnetic field H makes an angle ϕ with OP such that

$$\tan\phi=\frac{H_r}{H}=-\tan\frac{3\theta}{2},$$

i. e., H makes an angle $\frac{3\theta}{2}$ with the inward drawn radius vector PO.

(*d*) The radius of curvature of a pole piece, at the point where the axis Ox cuts it, is equal to twice the distance from the origin, *e. g.* the radius of curvature at E is equal to twice OE.

(*e*) In order to make the permeability μ of the pole pieces as large as possible so as to approximate to the condition $\mu=\infty$, it would seem better not to use such a large exciting current as would saturate the pole pieces magnetically, for then μ would have a less value than the maximum.

Magnetic Reluctance of the Air Gap.

The magnetic reluctance of the air gap between the pole pieces can be calculated if certain assumptions are made. Referring to fig. 1 let the air gap be assumed to be bounded by lines of force KSL, WVM, symmetrically situated relative to the line OH. Thus the edge of the pole piece on the left extends from K through O to W, and that on the right from L through G to M. The whole gap forms the figure OKLGMWO. It is proposed to calculate the reluctance per cm. depth measured normal to the plane of fig. 1 in the direction from WOK to LGM. The reluctance has the same form as the electric resistance, but the resistivity is to be replaced by unity since we are dealing with an air gap.

There is a theorem which states that the electric resistance of the medium between two perfect conductors measured in e.m.u. is equal to the specific resistance of the medium in e.m.u., divided by 4π times the electrostatic capacity of the same pair of conductors in e.s.u., regarded as the plates of a condenser with a vacuum between them. Consider then the condenser whose plates are KOW and LGM respectively. Let LGM be an equipotential $U=Cr^{\frac{3}{2}}\cos\frac{3\theta}{2}=K$, a constant. Then

$$r^{\frac{3}{2}}\cos\frac{3\theta}{2}=\frac{K}{C}=K_1^{\frac{3}{2}},$$

say. Any point along LGM has polar co-ordinates r and θ where $r=K_1\sec^{\frac{2}{3}}\frac{3\theta}{2}$. The charge σ per cm. depth at any point on LGM is $\frac{H}{4\pi}$ by Coulomb's theorem, regarding H as though it were the electrostatic field strength in e.s.u. Then Q_1, the total charge per cm. depth on $LGM=\int\sigma.ds$, where ds is an element of the curve. Hence $Q_1=\int\frac{H\,ds}{4\pi}$ along LGM

$$=2\int_{K_1}^{r}\frac{H}{4\pi}\left\{1+r^2\left(\frac{d\theta}{dr}\right)^2\right\}^{\frac{1}{2}}dr,$$

since $ds^2=dr^2+r^2d\theta^2$, the lower limit being $r=K_1$, its value at G, and the upper limit r, its value at L. The factor 2 is inserted because there is symmetry about G and the integral from M to G is equal to that from G to L. By the results of the last section $H=\frac{3Cr^{\frac{1}{2}}}{2}$. Hence

$$Q_1=2\int_{K_1}^{r}\frac{3Cr^{\frac{1}{2}}}{8\pi}\left\{1+r^2\left(\frac{d\theta}{dr}\right)^2\right\}^{\frac{1}{2}}dr.$$

Substituting

$$r=K_1\sec^{\frac{2}{3}}\frac{3\theta}{2}\quad\text{and}\quad\frac{dr}{d\theta}=K_1\sin\frac{3\theta}{2}\sec^{\frac{5}{3}}\frac{3\theta}{2},$$

it is found that

$$Q_1=\frac{3C}{4\pi}\int_{K_1}^{r}\frac{r^{\frac{1}{2}}dr}{\sin\frac{3\theta}{2}}=\frac{3C}{4\pi}\int_{K_1}^{r}\frac{r^2\,dr}{(r_3-K_1^{\,3})^{\frac{1}{2}}}.$$

Substituting $r^3=z$ this becomes, after integration,

$$Q_1=\frac{C}{2\pi}(r^3-K_1^{\,3})^{\frac{1}{2}}=\frac{C}{2\pi}K_1^{\frac{3}{2}}\tan\frac{3\theta}{2}.$$

Since the potential of $LGM=U=CK_1^{\frac{3}{2}}$, and the potential of KOW is $Cr^{\frac{3}{2}}\cos 90°=0$, the capacity per cm. depth is $\frac{Q_1}{U}=\frac{1}{2\pi}\tan\frac{3\theta}{2}$. Its reciprocal is $2\pi \cot \frac{3\theta}{2}$. Hence the reluctance per cm. depth$=R_1$ say,

$$=\frac{1}{4\pi . \text{capacity}}=\frac{1}{2}\cot\frac{3\theta}{2}.$$

Half the gap, viz. OKLGO, has twice this reluctance, *i. e.* $\cot\frac{3\theta}{2}$ itself. In these expressions θ is the angle subtended by the arc LG at O. Since reluctances can be treated like resistances and the regions OKSFO and FSLGF are air gaps in series, the reluctance of a portion like FSLGF per cm. depth is $\cot\frac{3\theta}{2}-\cot\frac{3\theta'}{2}$, where θ' is the angle subtended by SF at O. The total flux across the gap OKLGMWO is $4\pi Q_1$, where Q_1 is the total pole strength on LGM and is equal to

$$2Cr^{\frac{3}{2}}\sin\frac{3\theta}{2}=2K\tan\frac{3\theta}{2},$$

where r and θ are the polar co-ordinates of L. The above theory can easily be adapted to pole pieces of any shape given by $r^{\frac{3}{2}}\cos\frac{3\theta}{2}=\text{constant}$. Other examples are shown in figs. 4, A1, A2, B1, C1 and D1.

Measurement of Permeability.

The whole electromagnet with its yoke and air gap forms a magnetic circuit obeying the law

$$\text{Total Flux}=\frac{\text{Magnetomotive Force}}{\text{Reluctance}}.$$

This gives
$$2Cr^{\frac{3}{2}}\sin\frac{3\theta}{2}=\frac{1{\cdot}257Ni}{R_2+\frac{1}{2}\cot\frac{3\theta}{2}},$$

where Ni=the ampere turns and R_2 is the reluctance of the ferromagnetic yoke of the magnet. Hence

$$R_2=\frac{1{\cdot}257\,Ni}{2Cr^{\frac{3}{2}}\sin\frac{3\theta}{2}}-\frac{1}{2}\cot\frac{3\theta}{2}.$$

Since $\frac{9C^2}{4}$ is the value of $\frac{\partial H^2}{\partial r}$ along any line, including FG, it can be obtained from fluxmeter measurements of H along FG. This gives C; r and θ are known from the dimensions of the pole pieces. Hence R_2 can be calculated if N and i are known. Generally $R_2=\Sigma\frac{l}{\mu A}$, where l is the length, A the cross-section and μ the permeability of each portion of the yoke. If μ can be assumed to be constant throughout the yoke in any one experiment, $R_2=\frac{1}{\mu}\Sigma\frac{l}{A}$ and $\Sigma\frac{l}{A}$ can be calculated from the dimensions of the yoke. Hence from

$$\frac{1}{\mu}\Sigma\frac{l}{A}=\frac{1{\cdot}257\,Ni}{2Cr^{\frac{3}{2}}\sin\frac{3\theta}{2}}-\frac{1}{2}\cot\frac{3\theta}{2},$$

μ can be calculated. It will, of course, vary when i varies.

Other Magnetic Fields of Interest.

General case $$W=U+iV=Cz^n.$$

If, instead of using the index $\frac{3}{2}$ we use the general index n, we get the transformation

$$U+iV=Cz^n=C(re^{i\theta})^n=Cr^n\cos n\theta+iCr^n\sin n\theta.$$

Hence $$U=Cr^n\cos n\theta \quad\text{and}\quad V=Cr^n\sin n\theta.$$

In this case take $U=Cr^n\cos n\theta$ as the potential and calculate the component magnetic fields and mechanical forces per c.c. along and perpendicular to a radius vector OP at any point P.

The radial field $$H_r=-\frac{\partial U}{\partial r}=-Cnr^{n-1}\cos n\theta.$$

$$H_r^2=C^2n^2r^{2n-2}\cos^2 n\theta.$$

The perpendicular component

$$H_\theta=-\frac{1}{r}\frac{\partial U}{\partial\theta}=Cnr^{n-1}\sin n\theta.$$

$$H_\theta^2=C^2n^2r^{2n-2}\sin^2 n\theta.$$

$$H^2=H_r^2+H_\theta^2=C^2n^2r^{2n-2}.$$

Then $$\frac{\partial H^2}{\partial r}=C^2(2n-2)n^2r^{2n-3}=F_r \text{ say,}$$

and $$\frac{1}{r}\frac{\partial H^2}{\partial\theta}=0=F_\theta \text{ say.}$$

The resultant mechanical force per c.c. on a weakly magnetic body is $\frac{1}{2}(k-k_a)\frac{\partial H^2}{\partial s}$ dynes and

$$\frac{\partial H^2}{\partial s}=(F_r^2+F_\theta^2)^{\frac{1}{2}}=F_r \text{ since } F_\theta=0.$$

Hence $$\tfrac{1}{2}(k-k_a)\frac{\partial H^2}{\partial s}=\tfrac{1}{2}(k-k_a)C^2(2n-2)n^2r^{2n-3}.$$

This force is proportional to $(n-1)n^2r^{2n-3}$. It is equal to zero when $n=0$ or 1, and is constant when the index $2n-3=0$, *i. e.* when $n=\frac{3}{2}$. The case $n=0$ corresponds to U=constant and is of no special interest.

The case $n=1$ corresponds to $U=r\cos\theta$, $V=r\sin\theta$, *i. e.* $U=x$, $V=y$, and represents a uniform magnetic field, the lines of force being parallel to one axis of co-ordinates and the equipotentials to the other. The case $n=\frac{3}{2}$ is the one already discussed. On account of the factor $n-1$ it is seen that the force, in the general case, is directed away from the origin if $k>k_a$ and $n>1$, and towards the origin if $k>k_a$ and $n<1$, including negative values of n. H always makes an angle ϕ with the radius vector OP such that $\tan\phi=\frac{H_\theta}{H_r}=-\tan n\theta$. Since when n has been fixed r is the only variable, it is seen that except when $n=\frac{3}{2}$ the curves of constant force are given by putting $(n-1)n^2r^{2n-3}$ equal to a constant. This gives $r=$constant, so that the curves of constant force are circles with the origin as centre, in every case, except when $n=\frac{3}{2}$, when the force is constant everywhere. The force in every case is directed along the radius vector from the origin. No new information is obtained by exchanging the rôles of U and V, since the lines of force and the equipotentials are of the same shape, one set of curves being brought into coincidence with the other by a rotation through an angle $\frac{\pi}{2n}$. The magnetic reluctance per cm. depth of an air gap bounded on the left by a pole piece whose section has the equation $Cr^n\cos n\theta=0$ (two straight lines), and on the right by a pole piece whose section has the equation $Cr^n\cos n\theta=U=K$, a constant, is easily found by the same method as that used above in the special case when $n=\frac{3}{2}$. The lateral boundaries of the air gap are supposed to be lines of force $Cr^n\sin n\theta=$a constant, so that the pole pieces are of unequal face-width. The result of the calculation is that the capacity per cm. depth is $\frac{1}{2\pi}\tan n\theta$ e.s.u., and the magnetic reluctance $\frac{1}{2}\cot n\theta$ e.m.u. per cm. depth, a simple result considering that it represents a general case. Here θ is the angle subtended by half the section of the right-hand pole piece at the origin, *e. g.*, in fig. 1 it is the angle subtended by LG at O if LGM is the section of the right-hand pole piece. The total flux across the gap$=2K\tan n\theta$.

Field of a Single Circular Cylinder or Two Concentric Cylinders.

The properties of such a field in electrostatics are already well known, but, in passing, it may be worth while to consider it from the point of view of magnetism. Ordinary electrostatic theory, using Gauss's theorem, gives the field $H=\frac{2Q_1}{r}$ at any point in the air near a cylindrical pole piece, if Q_1 is the pole strength per unit length of the cylinder. Alternatively, if we use the method of conformal transformations, the appropriate one, as is well known in electrostatics, is

$$W=C\log z,$$

$$W=U+iV=C\log re^{i\theta},$$
$$=C\log r+iC\theta.$$

Taking U as the potential, $U=C\log r$ and the equipotentials U=constant are given by r=constant, *i. e.* they are concentric circular cylinders.

$$H_r=-\frac{\partial U}{\partial r}=-\frac{C}{r}\quad\text{and}\quad H_r^2=\frac{C^2}{r^2}.$$

$$H_\theta=-\frac{\partial U}{r\,\partial\theta}=0\quad\text{and}\quad H_\theta^2=0.$$

$$H^2=H_r^2+H_\theta^2=\frac{C^2}{r^2}\quad\text{and}\quad H=-\frac{C}{r}.$$

$$\frac{\partial H^2}{\partial r}=-\frac{2C^2}{r^3}=F_r\text{ say,}\quad\text{and}\quad\frac{1}{r}\cdot\frac{\partial H^2}{\partial\theta}=0=F_\theta\text{ say.}$$

Hence $$F=\{F_r^2+F_\theta^2\}^{\frac{1}{2}}=F_r=-\frac{2C^2}{r^3}.$$

The resultant force per c.c. on a small para- or diamagnetic body $=-\frac{1}{2}(k-k_a)\frac{2C^2}{r^3}$ dynes, *i. e.* is proportional to r^{-3}. This result also holds when the pole pieces are complete concentric cylinders, with an air gap between them, or sectors of them, or when the rôles of U and V are interchanged and the sections of faces of the pole pieces are complete or partial radii of sectors of circles. The reluctance of the complete air gap between pole pieces in the form of concentric circular cylinders is $2\log\frac{r_2}{r_1}$ e.m.u. per cm. depth, where r_1 and r_2 are the internal and external radii of the air gap, and the total flux across the gap is $2\pi C=4\pi Q_1$, where Q_1 is the total pole strength on either pole per cm. depth. This would seem to apply to the piece of apparatus called Hibbert's Standard Inductance and to the air gaps of the permanent magnets of moving coil galvanometers, whose pole pieces have faces of circular section, with a concentric cylinder of iron between them. Pole pieces in the form of arcs of concentric cylinders of radii r_1, r_2 have an air gap of reluctance $\frac{4\pi}{\theta}\log\frac{r_2}{r_1}$ e.m.u., where θ is the angle in radians subtended at the centre. Pole pieces in the form of portions of radii of concentric cylinders, *i. e.* of equal width r_2-r_1, have an air gap of reluctance $\frac{\theta}{\log\frac{r_2}{r_1}}$ e.m.u., where θ is the angle subtended by an arc of the air gap at the centre.

Pole Pieces in the form of a Convex Circular Cylinder facing a Plane or of Two Circular Cylinders.

Again using well-known results of electrostatics, the magnetic field in an air gap formed by a convex circular cylinder of a ferromagnetic facing a plane is given by the conformal transformation

$$W=U+iV=C\log\frac{z+a}{z-a}=C\log\frac{r_1e^{i\theta_1}}{r_2e^{i\theta_2}}=C\log\frac{r_1}{r_2}+iC(\theta_1-\theta_2),$$

where U=potential, V=the conjugate function and r_1, θ_1 are the polar co-ordinates of any point P relative to a point A$(-a, 0)$ on the x-axis, r_2, θ_2 the polar co-ordinates of P relative to a point B$(+a, 0)$ on the x-axis (see fig. 2). Hence $\mathrm{U}=\mathrm{C}\log\frac{r_1}{r_2}$ and $\mathrm{V}=\mathrm{C}(\theta_1-\theta_2)$. The equipotentials $\mathrm{U}=\mathrm{C}\log\frac{r_1}{r_2}=\mathrm{K}$ form a set of coaxial circles whose common axis is Ox, and whose centres are at such points as O′. The y-axis is the limiting form of one of these circles when $r_1=r_2$ and U=K=0. The lines of force are circles orthogonal to the equipotentials with Oy as a common axis, and they all pass through the points B and A. To calculate the mechanical force on a small para- or diamagnetic body the magnetic field H at P is required. Let the polar co-ordinates of P be r, θ relative to O. The well-known result obtained by differentiating U so as to get

$$\mathrm{H}_r=-\frac{\partial \mathrm{U}}{\partial r} \quad\text{and}\quad \mathrm{H}_\theta=-\frac{1}{r}\frac{\partial \mathrm{U}}{\partial \theta}$$

and then combining, is $\mathrm{H}=\frac{2\mathrm{C}a}{r_1r_2}$, where a=OB. Hence

$$\mathrm{H}^2=\frac{4\mathrm{C}^2a^2}{r_1{}^2r_2{}^2},\quad \textit{i. e.,}\quad \mathrm{H}^2=\frac{4\mathrm{C}^2a^2}{(a^2+r^2)^2-4a^2r^2\cos^2\theta}.$$

Then $$\mathrm{F}_r=\frac{\partial \mathrm{H}^2}{\partial r}=\frac{-16\mathrm{C}^2a^2r(a^2+r^2-2a^2\cos^2\theta)}{\{(a^2+r^2)^2-4a^2r^2\cos^2\theta\}^2}.$$

Also $$\mathrm{F}_\theta=\frac{1}{r}\cdot\frac{\partial \mathrm{H}^2}{\partial \theta}=\frac{-32\mathrm{C}^2a^4r\cos\theta\sin\theta}{\{(a^2+r^2)^2-4a^2r^2\cos^2\theta\}^2}.$$

$$\mathrm{F}=(\mathrm{F}_r{}^2+\mathrm{F}_\theta{}^2)^{\frac{1}{2}}=\frac{16\mathrm{C}^2a^2r}{\{(a^2+r^2)^2-4a^2r^2\cos^2\theta\}^{\frac{3}{2}}}.$$

The resultant mechanical force per c.c. on a small para- or diamagnetic body is $\frac{1}{2}(k-k_a)$F dynes. The force (or F) makes an angle ϕ with OP where $\tan\phi=\frac{\mathrm{F}_\theta}{\mathrm{F}_r}=\frac{a^2\sin 2\theta}{r^2-a^2\cos 2\theta}$, so that the differential equation of the curves of motion of the small body is $\frac{r\,d\theta}{dr}=\frac{a^2\sin 2\theta}{r^2-a^2\cos 2\theta}$. At points along O$x$, where $\theta=0$, or along a line θ=a very small quantity, $\mathrm{F}=\frac{16\mathrm{C}^2a^2r}{(a^2-r^2)^3}$. and along O$y$ where $\theta=90°$, $\mathrm{F}=\frac{16\mathrm{C}^2a^2r}{(a^2+r^2)^3}$. At points not far from O in any direction, *i. e.* when $r\ll a$, $\mathrm{F}=\frac{16\mathrm{C}^2r}{a^4}$ and $\tan\phi=-\tan 2\theta$. At points a great distance from the origin, *i. e.* when $r\gg a$, $\mathrm{F}=\frac{16\mathrm{C}^2a}{r^5}$ and $\tan\phi=0$ The curves of constant F, in the general case, are given by

$$\frac{16\mathrm{C}^2a^2r}{\{(a^2+r^2)^2-4a^2r^2\cos^2\theta\}^{\frac{3}{2}}}=\text{constant},$$

i. e., $\frac{r}{r_1{}^3r_2{}^3}$=constant or $r_1r_2=\mathrm{B}r^{\frac{1}{3}}$ where B is a constant. The magnetic reluctance per cm. depth of a symmetrical air gap OFEPMP′JGO, between FOG and PMP′, bounded laterally by the circular lines of force

FEP and GJP′, is proved by the same method as that used in an earlier case to be

$$\frac{\mathrm{K}}{4\tan^{-1}\left(\frac{\mathrm{K}+1}{\mathrm{K}-1}\tan\frac{\psi}{2}\right)}=\mathrm{R}_1 \text{ say, in e.m.u.,}$$

where ψ is the angle OO′P or OO′P′. Similarly the reluctance of the gap OFEIJGO from FOG to EIJ is

$$\frac{\mathrm{K}'}{4\tan^{-1}\left(\frac{\mathrm{K}'+1}{\mathrm{K}'-1}\tan\frac{\psi'}{2}\right)},$$

where K′ is the potential of EIJ and ψ' is the angle subtended by EI not at O′ but at its own centre of curvature. Hence the reluctance of a curvilinear air gap such as IEPMP′JI, from EIJ to PMP′, per cm. depth is

$$\frac{1}{4}\left\{\frac{\mathrm{K}}{\tan^{-1}\left(\frac{\mathrm{K}+1}{\mathrm{K}-1}\tan\frac{\psi}{2}\right)}-\frac{\mathrm{K}'}{\tan^{-1}\left(\frac{\mathrm{K}'+1}{\mathrm{K}'-1}\tan\frac{\psi'}{2}\right)}\right\}$$

e.m.u. For a gap such as CDHGP′MPFC, from PMP′ to CDH the reluctance per cm. depth is

$$\frac{1}{4}\left\{\frac{\mathrm{K}}{\tan^{-1}\left(\frac{\mathrm{K}+1}{\mathrm{K}-1}\tan\frac{\psi}{2}\right)}-\frac{\mathrm{K}''}{\tan^{-1}\left(\frac{\mathrm{K}''+1}{\mathrm{K}''-1}\tan\frac{\psi''}{2}\right)}\right\}$$

e.m.u., where K″ is the potential of CDH and ψ'' is the angle subtended by DC at O″. The minus sign arises from the negative value of the potential K″.

Experimental Tests.

The type of pole pieces now under discussion was employed by Bates Baker and Meakin (3), and the results shown in fig. 5, p. 433, of their paper can readily be explained on the basis of the foregoing treatment, but these were supplemented by the following confirmatory experiments.

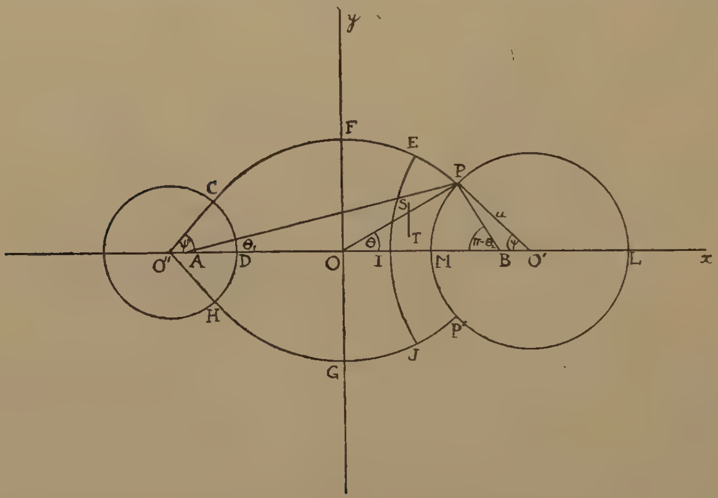
A plane pole piece facing a convex circular cylindrical pole piece was fixed on a Rutherford electromagnet, so as to form an air gap resembling OFEPMP′JGO in fig. 2. With the magnet excited, the magnetic field H at various points along OM$=\frac{2\mathrm{C}a}{r_1r_2}=\frac{2\mathrm{C}a}{(r+a)(r-a)}=\frac{2\mathrm{C}a}{r^2-a^2}$ in this case.

Re-arranging, this equation becomes $r^2-a^2=\frac{2\mathrm{C}a}{\mathrm{H}}$, and if we write $r^2=y$, $\frac{1}{\mathrm{H}}=x$, it becomes $y=2\mathrm{C}ax+a^2$. Numerical values of H were obtained by a fluxmeter method and a graph was plotted using $y=r^2$, $x=\frac{1}{\mathrm{H}}$. It proved to be a straight line, in accordance with $y=2\mathrm{C}ax+a^2$. Other plots such as $y=\mathrm{H}$, $x=r$ or $y=\mathrm{H}^2$, $x=r$ gave curved graphs. Numerical values of a and C were obtained from the straight-line graph. As another test a torsion balance was used. With a small volume of pure mercury suspended initially at a point such as T in fig. 2, and the axis of rotation at S, where S was on a normal to Ox through T, the magnetic field was switched on, causing a force to act on the mercury with a definite torque about the axis through S. When the arm of the balance had rotated through an angle E from its initial position ST towards the plane pole piece, theory predicted a value of the torque proportional to

$$\frac{r\cos(\mathrm{E}+\phi+\theta)}{\{a^4+r^4-2a^2r^2\cos 2\theta\}^{\frac{3}{2}}},$$

where ϕ=the angle between the force and the line from O to the body and the other symbols have their previous meanings. On plotting theoretical values of this expression as ordinates against values of E as abscissæ, a graph was obtained which fell away from a high initial value of the torque when E=0 to a value about 33 per cent. lower when E=8°. The theoretical graph was practically a straight line over this range, though slightly convex downwards over the longer range E=0° to 20°. The experimental value of the torque, in which errors up to ±5 per cent. were possible, fell off by about 37 per cent. over the range E=0° to 8°, and gave a straight-line graph over this range.

Fig. 2.



$\angle$PO'B$=\psi$. $\angle$CO''D$=\psi''$.
OA=OB=a.
O'P=O'L=u.
OP=r. $\angle$POM$=\theta$.
AP=r_1. $\angle$PAO$=\theta_1$.
PB=r_2. $\angle$PBO'$=\theta_2$.

Concave Cylindrical Pole Piece facing a Plane, Convex or Concave One.

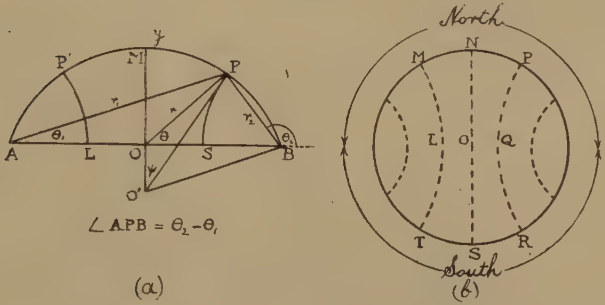
The same transformation

$$\mathrm{U}+i\mathrm{V}=\mathrm{C}\log\frac{r_1e^{i\theta_1}}{r_2e^{i\theta_2}}=\mathrm{C}\log\frac{r_1}{r_2}+i\mathrm{C}(\theta_1-\theta_2),$$

solves the problem of an air gap between a concave circular cylindrical pole piece and a plane one or between two concave cylindrical ones. We have simply to exchange the rôles of U and V and to take $V=C(\theta_1-\theta_2)$ as the potential. Thus, in fig. 3 (*a*), let the circular arc AMB represent an equipotential $V=C(\theta_1-\theta_2)=CK$, and the straight line AB the equipotential $V=-C\pi$, since $\theta_1-\theta_2=-\pi$ along AB. The lines of force are the set of orthogonal circles which were equipotentials before. As before, A has co-ordinates $-a$, 0 and B $+a$, 0. All equipotentials pass through A and B. Just as before, the field-strength H at any point P is $\dfrac{2aC}{r_1r_2}$. F, the variable part of the mechanical force on a small body, is

$$\frac{16C^2a^2r}{\{(a^2+r^2)^2-4a^2r^2\cos^2\theta\}^{\frac{3}{2}}}$$

Fig. 3.



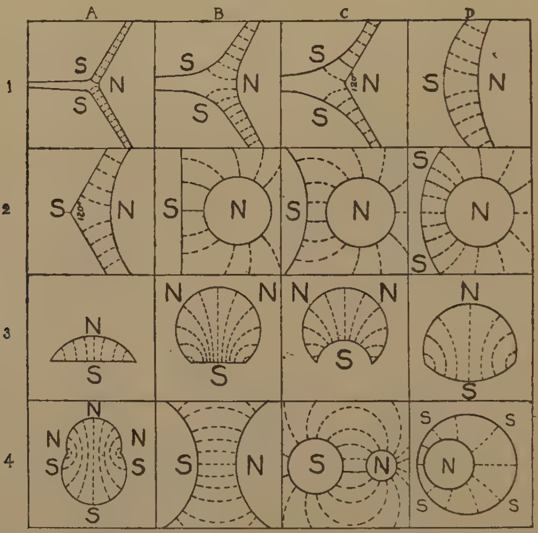
(a) (b)

dynes per c.c. and the actual force is $\frac{1}{2}(k-k_a)F$ dynes per c.c. When real pole pieces are made after this shape there should be no lateral edge effect of "fringing field" if the concave pole piece is brought right round to all but meet the plane one at A and B, but, of course, the edge effect due to the finite nature of the pole pieces in the vertical direction (top and bottom) will still exist. The reluctance of the complete air gap from AMB to AOB in fig. 3 (*a*) is zero, because there is ferromagnetic contact at the ends and the rest of the gap is in parallel with the infinitely short end portions. The reluctance of the symmetrical air gap PSOLP′MP between P′MP and LOS is shown by the same method as that used before to be

$$\frac{(\pi+K)}{2\log\dfrac{B+a\tan\frac{\psi}{2}}{B-a\tan\frac{\psi}{2}}}$$

e.m.u. per cm. depth, where B=OM, a=OA or OB, and ψ is the angle subtended by the arc PM at O′, the centre of curvature of PM. K is π minus the angle which AB subtends at any point along PMP′. If we call this reluctance R_1, the reluctance of half the gap, *i. e.* of MPSOM between MP and SO, is $2R_1$. If the air gap is that between equal semi-cylinders, *e. g.* MNPQRSTLM in fig. 3 (*b*), the reluctance between MNP and TSR is again R_1. That between NP and SR is $2R_1$. In all these cases the lateral boundaries of the air gap are circular lines of force such as PQR and MLT. If the pole pieces are cylinders of different radii, *e. g.* like those represented in fig. 4, C 3, D 3 and A 4, the total reluctance of the air gap can be obtained by adding together the reluctances of the parts on the two sides of the plane LOS fig. 3 (*a*). Fig. 4, from B 2 to D 4, represents examples of convex, plane and concave pole pieces facing each other, I am aware that, in a certain sense, all these figures are included in a well-known diagram which perhaps first appeared in the plate accompanying a memoir by Messrs. Foster and Lodge in the very first volume

Fig. 4.



The squares are identified somewhat as in maps. A 1, A 2, B 1, C 1, D1 are pole pieces and air gaps giving a constant mechanical force on a small body.

B2 to D4, inclusive, are air gaps of circular cylindrical, convex, concave and plane pole pieces.

In figs. A 3, A 4, B 3, C 3 and D 3, the two pole pieces must be supposed not to make contact with each other, or else the air gap will be shorted and there will be no magnetic field in it. Hence the reader should assume that a very narrow gap separates the pole pieces near the points of discontinuity at the sides.

of the ‘ Proceedings of the Physical Society,’ concerning the corresponding problems in current electricity. From the point of view of the design of magnet pole pieces, however, clarity is gained by showing separate diagrams, with the lines of force distinguished from the faces of the pole pieces. It is important to note that in all these air gaps or in other derived gaps bounded laterally by any pair of lines of force shown, and terminated by the corresponding portions of the pole pieces, the magnetic field at a point has the value $H=\frac{2aC}{r_1 r_2}$ and the mechanical force per c.c. on a small para- or diamagnetic body has the value

$$\tfrac{1}{2}(k-k_a)F=\frac{8(k-k_a)C^2a^2r}{r_1{}^3r_2{}^3}\text{ dynes,}$$

at points with the same coordinates relative to O, A and B in figs. 2, 3 *a*, 3 *b* and 4, B 2 to D 4, inclusive.

Mechanical Force on a Small Specimen near a Short Bar Magnet.

It is proposed to calculate the mechanical force per c.c. experienced by a small para- or diagmagnetic body placed at any point P near a short bar magnet, and also to find the equation to the curves of motion which such a body would describe, if free to move. It is important to distinguish this force from the magnetic field strength, which is the force on a single (particle) N. pole, and to distinguish the curves of motion of this body from the lines of force.

We will calculate the factor $\frac{\partial H^2}{\partial s}$, using fig. 5. The potential U at P is $\frac{M\cos\theta}{r^2}$ by the usual theory. Let the component magnetic fields at P, along and normal to OP be H_r and H_θ. Then

$$H_r=\frac{2M\cos\theta}{r^3},\quad H_r{}^2=\frac{4M^2\cos^2\theta}{r^6},\quad \frac{\partial H_r{}^2}{\partial r}=-\frac{24M^2\cos^2\theta}{r^7}.$$

$$H_\theta=\frac{M\sin\theta}{r^3},\quad H_\theta{}^2=\frac{M^2\sin^2\theta}{r^6},\quad \frac{\partial H_\theta{}^2}{\partial r}=-\frac{6M^2\sin^2\theta}{r^7}.$$

F_r, the component of $\frac{\partial H^2}{\partial s}$ along OP,

$$=\frac{\partial H_r{}^2}{\partial r}+\frac{\partial H_\theta{}^2}{\partial r}=\frac{-6M^2(\sin^2\theta+4\cos^2\theta)}{r^7},$$

F_θ, the component of $\frac{\partial H^2}{\partial s}$ normal to OP,

$$=\frac{1}{r}\frac{\partial H_r{}^2}{\partial\theta}+\frac{1}{r}\frac{\partial H_\theta{}^2}{\partial r}=\frac{-6M^2\cos\theta\sin\theta}{r^7},$$

since $\frac{1}{r}\frac{\partial H_r{}^2}{\partial\theta}=\frac{-8M^2\cos\theta\sin\theta}{r^7}$ and $\frac{1}{r}\frac{\partial H\ {}^2}{\partial\theta}=\frac{+2M^2\cos\theta\sin\theta}{r^7}$.

$\frac{\partial H^2}{\partial s}$ itself is the square root of the sum of the squares of these components, *i. e.*,

$$\frac{\partial H^2}{\partial s}=\frac{6M^2(\sin^4\theta+9\sin^2\theta\cos^2\theta+16\cos^4\theta)^{\frac{1}{2}}}{r^7},$$

and the actual force per c.c. is this quantity multiplied by $\tfrac{1}{2}(k-k_a)$.

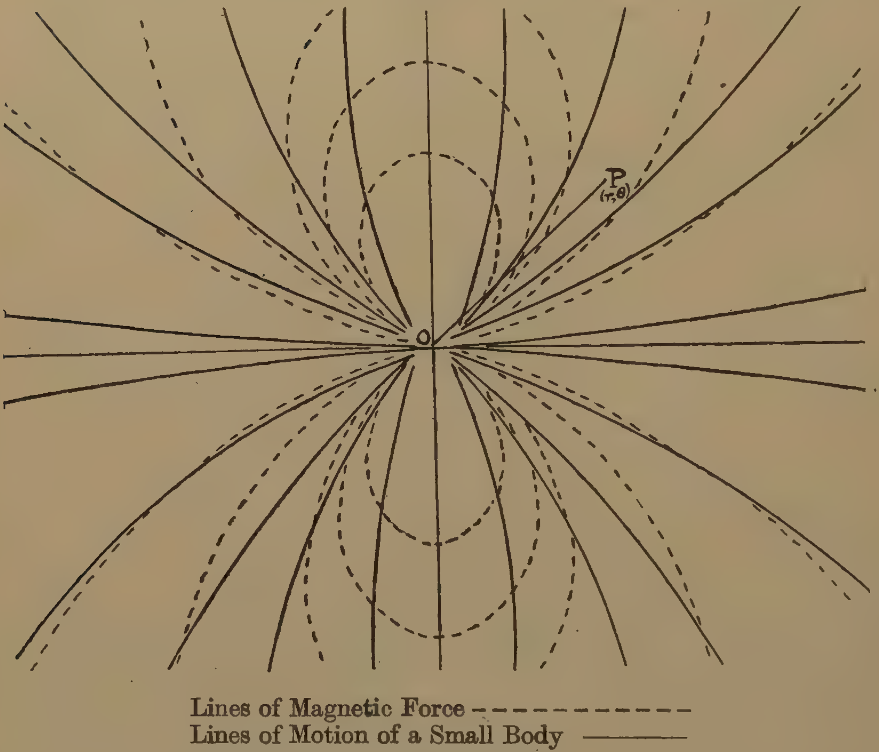
It makes an angle ϕ with OP such that $\tan\phi=\frac{F_\theta}{F_r}$. The lines along which the small body would move, if free, are given by putting

$$\frac{dr}{r\,d\theta}=\frac{F_r}{F_\theta}=\frac{\sin^2\theta+4\cos^2\theta}{\sin\theta\cos\theta}.$$

Hence $$\frac{dr}{r}=\tan\theta\,.\,d\theta+4\cot\theta\,.\,d\theta.$$

Integrating, $$\log r=\text{constant}-\log\cos\theta+4\log\sin\theta.$$

Fig. 5.


P
(r,θ)
O
Lines of Magnetic Force - - - - - - - - - - - -
Lines of Motion of a Small Body ————

Hence $r=\frac{A\sin^4\theta}{\cos\theta}$ is the equation to a curve of motion. We may compare it with $r=B\sin^2\theta$, the polar equation to a line of force. Fig. 5 represents the two sets of curves on the same diagram. It appears that a small para- or diamagnetic body would not, except when $\theta=0$, move along a line of force.

The case of a ferromagnetic body is complicated by the presence of an extra tensile force $2\pi(k-k_a)H_2^2\cos^2\theta_2$ per sq. cm. along the outward normal. If, however, the body is a small sphere and is placed where there is no excessive curvature of lines of force, the resultant tensile force due to this term should be zero due to equal and opposite pulls on opposite sides of the sphere. Of course, the ferromagnetic will distort the field in its neighbourhood. On trial by experiment, a small initially unmagnetized iron sphere floating on a cork in water near a short straight electromagnet also moved along curves resembling the curves of motion calculated above and shown in fig. 5. In ferromagnetics, k is itself a function of H.

At a point along the extended axis of a short bar magnet $\frac{\partial H^2}{\partial s}=\frac{24M^2}{r^7}$, and at a point along an equatorial line $\frac{\partial H^2}{\partial s}=\frac{6M^2}{r^7}$, so that the ratio of the forces per c.c. acting on a small para- or diamagnetic body placed at equal distances in these two positions is as four to one. If the law of magnetic force had been $F=\frac{m_1m_2}{r^n}$ the ratio would have been as n^2 to one. This suggests a torsion balance experiment to check or confirm the law of inverse squares.

The curves of constant force near a short bar magnet are given by putting $\frac{\partial H^2}{\partial s}$ equal to a constant.

This gives $$r^7=B(\sin^4\theta+9\sin^2\theta\cos^2\theta+16\cos^4\theta)^{\frac{1}{2}}$$

or $$r^{14}=C(\sin^4\theta+9\sin^2\theta\cos^2\theta+16\cos^4\theta).$$

A law of force obeying an inverse seventh-power law of distance implies a very rapid decrease in force as we move away from the magnet. Assuming the force at a point on the axis to be $\frac{1}{2}(k-k_a)\frac{\partial H^2}{\partial r}$ per c.c., this is equal to 120/128 dyne per c.c. when $M=1{,}000$ e.m.u., $r=2$ cm., $(k-k_a)=10^{-5}$ e.m.u.

I desire to thank Professor L. F. Bates for placing the resources of his research laboratory at my disposal for the experimental tests of the above theory, to Mr. E. M. Somekh, B.Sc., for carrying out those tests, and to Mr. F. Underwood, M.Sc., for checking the first part of the theory.

References.

(1) J. J. Thomson, ‘Elements of the Mathematical Theory of Electricity and Magnetism,’ 5th ed. p. 290.

(2) Fereday, Proc. Phys. Soc. xlii. p. 251 (1930), and Proc. Phys. Soc. xliii. p. 383 (1931).

(3) Bates, Baker and Meakin, Proc. Phys. Soc. lii. p. 425 (1940).

LXII. *Contact Potentials.*—Part V. *The Metal-Electrolyte Volta Effect.*

By J. Alan Chalmers, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham *.

[Received March 20, 1942.]

1. *Introduction.*

In earlier parts, general principles have been laid down for the discussion of contact potentials and these have been applied to the discussion of the Volta effect for metals and the Voltaic cell, particular attention being drawn to the concentration cell.

In the case of the metal-electrolyte Volta effect, there are three junctions ; of these, the metal-air junction has already been discussed in connection with the Volta effect, but it will be convenient to discuss it in a slightly different way in the present case ; the metal-electrolyte junction has been discussed in connection with the Voltaic cell ; and the electrolyte-air junction requires special treatment.

In discussing the Volta effect, the carrier of charge within the dielectric has been considered to be an electron, but now it will be convenient to consider, instead, that it is a metal ion.

It is impossible to consider the dielectric to be a vacuum, but, as has been shown in Part II., results for a dielectric, such as air, in place of a vacuum show no difference unless chemical effects occur.

As in the case of the ordinary Volta effect, and perhaps more so, surface layers affect the metal-electrolyte Volta effect.

2. *The Metal-Air Junction.*

The total potential difference of a system containing a dielectric can be measured by the work done in taking a unit charge across the dielectric ; in Part II., the discussion of the Volta effect was carried out considering that the charge carried across was composed of electrons, but the result will be the same whatever carries the charge, and it is here convenient to consider the charge to be carried by metal ions.

At the metal-air junction, if metal ions are considered to take the current across the junction, there will be a change of carrier potential difference ${}_3C_1$ since D_1 (for electrons) differs from ${}_3D_1$ (for metal ions) ; and the transfer potential difference, M_{13} is now ${}_3D_1$. The external potential difference must remain zero, for there is no change of potential as between a metal ion outside the metal and " outside the air." And it can easily be seen that the internal potential difference remains D_1, the terms ${}_3D_1$ in ${}_3C_1$ and in M_{13} cancelling out. So the consideration of the metal ion as the carrier, in place of the electron, makes no difference either to the external or internal potential differences, and the previous results for the metal-air junction still hold.

---

* Communicated by the Author.

3. *The Electrolyte-Air Junction.*

The results of Part I. give the internal potential difference

$$B_{23}=D_2-D_3+A_{23}.$$

Now D_3 is zero, since there is no change of potential in removing any carrier from the air; D_2 is given, from Part III., as $\frac{v_2^-P_2^-+v_2^+P_2^+}{v_2^-+v_2^+}$, and so all that is needed to obtain B_{23} is the knowledge of A_{23}.

It might be expected that A_{23} would be zero, just as A_{13} is zero at the metal-air junction, but there are reasons, which will be discussed later, for doubting this, and, for the present, the term A_{23} will be retained.

4. *The Total Potential Difference of the System.*

Considering external potential differences, we see that the sum will be $A_{12}+A_{23}$, A_{13} being zero. $A_{12}={}_1D_2-{}_2D_1=P_2^+-{}_2D_1$. Clearly, the same result is obtained by considering internal potential differences.

If A_{23} is zero, then the total potential difference of the system is equal to A_{12}, and this fact has been used by various workers in attempts to obtain A_{12}, the "single potential difference" at a metal-electrolyte junction. Early attempts of this sort, *e. g.* by Exner and Tuma (1), neglected various other boundaries that also existed in the system used, but Andauer (2) and others have made attempts free from such objections. In Andauer's work, the effect of surface layers at the metal-air boundary appeared to vitiate most of the results, but later work, particularly by Klein and Lange (3), overcame these difficulties and would appear to give reasonably accurate results for $A_{12}+A_{23}$, and so, if A_{23} is zero, for A_{12}.

The effect of surface layers at the electrolyte-air boundary has been investigated by Schulman and Rideal (4) and others, who have been able to obtain information about the properties of thin films.

5. *Surface Layers.*

From the general results obtained in Part I., we see that surface layers at the metal-electrolyte boundary do not affect the total external, or internal, potential difference at this boundary, but that surface layers at the metal-air boundary do affect potential differences there. Some of the methods used for determining "single potential differences" at a metal-electrolyte boundary do, in fact, measure external potential differences at the boundary of relative motion; thus, if a surface layer is present, what is measured is the external potential difference at the boundary of the surface layer and the electrolyte. In adding the external potential differences of the complete circuit, we will have two metal-surface layer junctions which will annul one another provided the two surface layers are the same, one at the metal-air junction and the other at the metal-electrolyte junction; in such a case, the metal-electrolyte Volta effect will measure the external potential difference at the surface layer-electrolyte junction together with A_{23} at the electrolyte-air junction.

6. *The Electrolyte-Air External Potential Difference.*

In discussing the potential differences connected with the electrolyte, we have considered the change of potential in taking an ion from the electrolyte to "just outside"; if this is the same at the electrolyte-air junction as it is at the metal-electrolyte junction, then we can have no external potential difference at the former. But if, owing to the nature of the water-air surface, it involves a different change of potential to remove a metal ion at the different places, then we can conveniently consider the metal ion to be "just outside" when there has been the same change of potential as moves it to just outside at the metal-electrolyte junction, and then consider the ion to have to move through an additional potential difference (A_{23}) at the electrolyte-air boundary. Alternatively, we can consider that there is a surface layer in the electrolyte at the electrolyte-air boundary which does not exist at the metal-electrolyte boundary; there is no external potential difference between the surface layer and the air, but there may be one (A_{23}) between the "pure" electrolyte and the surface layer. It would, of course, be possible to consider the position of "just outside" to be outside the electrolyte at the electrolyte-air boundary, and then to consider a surface layer at the metal-electrolyte boundary; this is less convenient for discussion than the other way of discussing the problem for it would require reconsideration of the metal-electrolyte boundary, and an alteration in the value of P_2^+.

No direct method is available for the determination of the value of the external potential difference at the electrolyte-air boundary, for all systems which can give a measured potential difference contain at least three junctions and there is no certain means of determining some of these independently of the others.

7. *Experiments with Filter Paper, etc.*

Chalmers and Pasquill [(5)] carried out a series of experiments, in which they found the external potential difference of a system of copper/moist filter paper/air/water/copper, obtaining a definite value of −0·26 volts; when the moist filter paper was replaced by various other absorbers of water, *e. g.* cotton wool, linen, $BaSO_4$ paste, etc., the potential difference of the system was unaltered, but when it was replaced by running water or a stationary water film, then, as would be expected, the result was zero. The effect was unaltered by the replacement of copper by zinc, or by alteration in other conditions of the experiment.

The system filter paper/copper/water gives zero potential difference, as can be seen by the consideration of the effect of taking a charge round the circuit. Thus we see that the result of −0·26 volts must occur in the system filter paper/air/water, and we must obtain this potential difference as the sum of the external potential differences filter paper/air,

air/water and water/filter paper. Of these, the last one also occurs in the system filter paper/copper/water which has zero potential difference, and in which it seems reasonable to assume that this junction has zero external potential difference ; if this is so, then the −0·26 volts remains at the filter paper/air and air/water boundaries.

Chalmers and Pasquill have suggested that any additional potential difference at the air-water boundary will be due to the orientation of the water molecules at such a boundary. In the case of filter paper, it is the fibres of the filter paper which govern the orientation of the molecules and we can thus expect a purely random orientation, which will give no external potential difference. This leaves the 0·26 volts to be supplied by the water/air boundary, and due to the orientation of water molecules by the presence of the air, an orientation which is not likely to occur in the same way at a metal-water or filter paper-water boundary. Although further confirmation would be necessary before it would be possible to state definitely that the potential difference of 0·26 volts is to be ascribed to the water/air junction, and not, wholly or in part, to the other two junctions in the system, it seems most reasonable to ascribe it to the water-air junction.

The measurements of Chalmers and Pasquill were carried out only with water, and not with electrolytes, but there is good reason to believe that the presence of an electrolyte will have no effect. Klein and Lange (3) have measured the Volta effect for silver and silver nitrate solutions of various concentrations, and have found the variation of the potential difference to be just that to be expected from the effect on the external potential difference at the metal-electrolyte boundary, showing that other potential differences in the system must be unaffected by the change of concentration. So we should expect the electrolyte-air external potential difference to be 0·26 volts whatever ions are dissolved in the electrolyte.

8. *Silver Measurements.*

If there is an electrolyte-air potential difference of 0·26 volts, then the metal-electrolyte Volta effect potential difference must be equal to this together with the external potential difference at the metal-electrolyte junction ; if this last can be measured independently, it will be possible to verify the value of 0·26 volts for the electrolyte-air potential difference.

A number of measurements of the external potential difference at the metal-electrolyte junction have been made with mercury as the metal, but, for reasons which will be discussed in Part VI., on " single potential differences," there is some doubt as to whether these measurements give all the external potential differences at the junction. Measurements have also been made at silver-electrolyte junctions, and these are more likely to give the required conditions. The results show that a junction of silver and normal silver nitrate has an external potential difference of about 0·25 volts, the silver being positive. Klein and Lange (3) have obtained the Volta effect for silver and normal silver nitrate and have found the result of 0·53 volts, the silver being positive, assuming the mercury-silver potential difference from photoelectric data ; this would agree with a value of about 0·28 volts for the electrolyte-air potential difference, which does not differ greatly from the value of 0·26 volts.

Perhaps an even better way of seeing this conclusion is to make use of the fact that the silver-silver nitrate external potential difference becomes zero when the concentration of the solution is about 5×10^{-5} normal. Considering the Volta effect of silver and this " null " solution, and extrapolating to this concentration the experimental results of Klein and Lange (3), the Volta potential difference is about 0·25 volts. There is no external potential difference at either the metal-air or metal-electrolyte junction, and so the only origin for the 0·25 volts potential difference must be the electrolyte-air boundary.

This argument appears to confirm the previous results as to the existence of a potential difference at the electrolyte-air junction.

9. *Dielectric Boundaries.*

In Part II. we have seen that there will be an external potential difference at a boundary of two dielectrics, depending on the two dielectric constants. It may therefore be suggested that there will be such an external potential difference at air-water and air-filter paper boundaries, even though the water and filter paper act as conductors of electricity ; but we cannot extend this idea to the question of the water-filter paper boundary since here there is equilibrium of ions and so no question of the same conditions as at a dielectric boundary. If the system could be considered as a system of three dielectrics, there would be no total potential difference, as discussed in Part II., but we can get a total potential difference if there are two dielectric boundaries and one electrolyte boundary. In this case, we should expect the external potential difference to be concentrated at the water-air boundary where the difference in dielectric constant is 79, rather than at the filter paper-air boundary where it is very much smaller ; so we get confirmation of the idea of an external potential difference of about 0·26 volts at the water-air boundary, rather than at either of the other junctions.

10. *Conclusion.*

The metal-electrolyte Volta effect has been discussed, and there seems to be good reason for believing that, in addition to the external potential difference at the metal-electrolyte boundary, there is another, of magnitude 0·26 volts, at the electrolyte-air boundary, possibly due to the orientation of the water molecules.

References.

(1) Exner and Tuma, *Sitz. Kais. Acad. Wien.* xcvii. p. 917 ; c. p. 607 ; ci. pp. 627, 1426 (1889–92).
(2) Andauer, *Zeit. Phys. Chem.* cxxv. p. 135 (1927).
(3) Klein and Lange, *Zeit. für Elektrochem.* xliii. p. 570 (1937).
(4) Schulman and Rideal, Proc. Roy. Soc. cxxx. p. 259 (1931).
(5) Chalmers and Pasquill, Phil. Mag. xxiii. p. 88 (1937).

LXIII. *Contact Potentials.*—Part VI. *Single Potential Differences.*

By J. ALAN CHALMERS, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham *.

[Received March 20, 1942.]

1. *Introduction.*

IN previous parts we have discussed definitions of potential differences at phase boundaries, and have applied the ideas to the Volta effect and the Voltaic cell. In this part we shall discuss the attempts that have been made to measure " single potential differences " at metal-electrolyte boundaries, and we shall endeavour to give a consistent explanation of the results obtained.

Numerous measurements have been made of the conditions under which a " potential difference " at a metal-electrolyte boundary can be annulled, and thence it is often possible to deduce the potential difference at the boundary under normal conditions. It is probable that, in all these cases, the quantity measured is what we have called the external potential difference at the boundary concerned. Because the normal calomel electrode, consisting of mercury, mercurous chloride paste and normal KCl solution, has been used as a reference standard of potentials, it seemed desirable to get an " absolute " value for the " electrode potential " of the normal calomel electrode ; therefore measurements have been made of the total potential differences of cells containing a normal calomel electrode, a salt bridge, and a boundary at which the potential difference has been measured, or in some cases annulled.

In some cases, *e. g.* when the boundary was that of an electrolyte and a metal other than mercury, the metal-metal junction should have been considered, but often was not. The results are thus subject to the same considerations as we have already discussed in relation to " electrode potentials " in Part III. When the measurements have been made with mercury as the metal, there is no metal-metal boundary to consider, and it would then seem that the only external potential differences in the circuit are those at the mercury-electrolyte junction and at the calomel electrode ; if the measurement made is that of the whole external potential difference at the mercury-electrolyte junction, then the results will certainly give the external potential difference at the calomel electrode ; but, as we shall discuss later, it is possible that there is another boundary at the mercury-electrolyte junction, also having an external potential difference.

---

* Communicated by the Author.

2. *The Capillary Electrometer and Similar Measurements.*

The first attempt to find the single potential difference at an electrode was that of the capillary electrometer, due to Lippmann (1). Essentially, this involves a method of measuring the interfacial surface tension at a boundary of mercury and acid, or salt solution, with various potential differences applied across the boundary; actually, in the circuit there are two mercury-acid boundaries, but, since one has a much smaller area of contact, and so much smaller capacity, the applied potential is almost entirely effective across this boundary. It is found that there is a definite potential difference at which the interfacial surface tension is a maximum. If there is a double layer of electric charge at the boundary, the magnitudes of the charges will depend on the potential difference at the boundary, composed of the natural external potential difference together with the applied potential difference. Lippmann suggested the law: $\gamma=\gamma_0-\alpha\sigma^2$, where γ is the surface tension when the double layer carries charges $\pm\sigma$, γ_0 and α being constant; it follows that γ will be a maximum when σ is zero, *i. e.*, when the applied potential difference just balances the natural external potential difference. The measured potential difference applied when the surface tension is a maximum is thus equal and opposite to the external potential difference at the mercury-acid junction. If we now measure the total potential difference of a cell consisting of such a mercury-acid junction, a salt bridge and a normal calomel electrode, we find the potential of the latter to be 0·56 volts. [In measuring single potential differences, the sign is taken as positive if the metal is positively charged.]

Kucera (2) and Craxford (3) have measured the change of interfacial surface tension by weighing drops of mercury in acid, with different applied potentials; the results agree entirely with those of the capillary electrometer.

A different method has been used by Palmaer (4) to give the potential difference at a mercury-acid junction; a jet of mercury is allowed to break up into drops just inside the acid, and a measurement is made of the potential difference between the jet and a stationary mercury electrode within the acid. If the time during which the mercury in the jet is in contact both with the acid and with the rest of the jet is sufficiently small, *i. e.*, if the jet breaks up very soon after entering the acid, there will not be time for the potential difference at the mercury-acid surface to be set up, and so the jet will be at the same potential as the acid. Hence the measured potential difference of the circuit is that of the stationary mercury surface. The results obtained agree with those obtained with the capillary electrometer.

Palmaer (5) has also used a slightly different method; if KCN or H_2S is added to the acid, the potential difference of the dropping electrode is reduced and, by a suitable addition, a " null " solution can be obtained, in which there is no potential difference between the dropping and stationary electrodes; then there will be no potential difference at the stationary electrode and a cell composed of mercury, null solution, salt bridge and calomel electrode would have only the potential difference of the calomel electrode, which is again found to be 0·56 volts.

It would seem that this result is independent of the assumption, in the capillary electrometer results, of the variation of surface tension with charge or potential difference ; but Heyrovsky (6) has pointed out that, if there is some charge or potential difference not necessarily zero, at a boundary of maximum surface tension, then, on the production of a drop from a jet, there must be set up a potential difference of this magnitude in order to comply with Le Chatelier's principle, and so the results for the dropping electrode must agree with those for the capillary electrometer, whether or not they measure the actual potential difference at the boundary.

3. *The Charges on Mercury Drops.*

Heyrovsky's argument does not apply to the question of the charges on mercury drops which have had time to reach electrical equilibrium with the solution. Christiansen (7) found that mercury drops falling in acid are deflected by a transverse field, carrying a positive charge. Frumkin (8) obtained a null solution, in which the drops are undeflected ; completing a circuit with a salt bridge and a calomel electrode gave the same result of 0·56 volts as in the previously-described experiments, showing that the condition of zero charge does agree with the condition of maximum surface tension.

Using a drop on a horizontal plate, Bordforss (9) and Craxford (40) found a null solution to prevent motion in a field, and Craxford also found the necessary applied potential for this. The results again agree with those previously discussed.

4. *The Mercury-Solution Boundary.*

There are some results in apparent conflict with those previously described. Bennewitz and Delijannis (11) and Bennewitz and Küchler (12) found that the results for the capillary electrometer with mercury in pure mercurous salts gave a result for the calomel electrode of about —0·2 volts instead of the usual value of 0·56 volts. For consistency with results which will be discussed later, we shall take a value of —0·15 volts instead of —0·20 volts. Craxford (10) verified these results, but found that they only occurred with pure mercurous salts, whereas if there is an acid or a salt of sodium or potassium the usual results occur. It has been considered by Craxford and others that the results with mercurous salts alone were vitiated by secondary effects and that the capillary electrometer and similar experiments give the current value for the external potential difference of the calomel electrode. But there are two other points for consideration ; in the first place, there are the results for the Volta effect which will be considered later. And there is the question, mentioned in Part III., as to how a mercury drop or surface can acquire a positive charge in a solution in which there can be very few positive ions of mercury. Even in a solution containing few mercury ions, mercury drops get a positive charge, and this would seem to be too great for it to be obtained from the mercury ions ; the positive charge must surely come from the other ions, hydrogen, sodium or potassium. If this is so, then there will be a layer on the mercury containing hydrogen, sodium or potassium, perhaps in the form of a kind of amalgam ; thus we would have at the mercury-solution boundary two phase boundaries, one between mercury and amalgam and the other between amalgam and solution. We can imagine that, whenever a mercury surface is in contact with a solution containing hydrogen, sodium or potassium ions, then some of these ions cross the boundary to form the " amalgam," and there will be an external potential difference between amalgam and solution, the amalgam being positive ; there is also an external potential difference between the amalgam and the pure mercury, and the total external potential difference at the mercury-solution boundary, which is what is required in discussing the potential difference of a cell, differs from the external potential difference at the amalgam-solution boundary, which is what is measured or annulled in experiments such as that of the capillary electrometer. Denoting by 1′ the amalgam, then the capillary electrometer measures $A_{1'3}$ but the cell includes also $A_{11'}$.

If we consider the amalgam to contain a constant number of ions, then the passage of a current through the boundary will only transfer mercury ions and not hydrogen, sodium or potassium ions, so we see that $A_{1'3}$, though set up by these ions, depends on the concentration of mercury ions in the solution. If we consider a system of pure mercury, amalgam and solution, we see that this has zero potential difference, since a current round will just carry mercury round, thus giving no chemical change, and no concentration change is possible. So we see that $A_{11'}+A_{1'3}=A_{13}$, which is the external potential difference which would occur, for the same concentration of mercurous ions, if the amalgam were absent.

A cell composed of a calomel electrode, a salt bridge, a mercurous salt and mercury gives a total potential difference amounting to 0·52 volts (taken as 0·50 volts approximately) for unit activity of the mercurous ions and independent of a presence of other ions. When mercurous salts alone are present, there is only a single boundary at the mercury-solution junction, and this has an external potential difference of 0·35 volts for unit activity of the mercurous ions, giving the external potential difference of the calomel electrode as —0·15 volts. When the " amalgam" is present, there is an external potential difference at the mercury-amalgam boundary of —0·70 volts and at the amalgam-solution boundary of 1·05 volts for unit activity.

Confirmation of this interpretation can be obtained from measurements of the Volta effect.

5. *Volta Effect Results.*

Measurements of the metal-electrolyte Volta effect have been made by Klein and Lange (13) in a way which yields consistent values. As we have discussed in Part V., these measurements give the external potential difference at the junction concerned provided that allowance is made for the external potential difference at the air-water interface, measured as 0·26 volts by Chalmers and Pasquill (14), and deduced from the silver results in Part V. as about 0·25 volts. Klein and Lange found the Volta effect for the mercury-normal solution junction to be 0·69 volts, when argon was used as the gas ; for moist air, their results are about 0·08 volts lower ; using a value of 0·60 volts and using 0·25 volts for the air-water external potential difference, we arrive at a value of 0·35 volts for the mercury-normal solution potential difference, in good agreement with the results for mercurous salts in solution, but quite at variance with the ordinary capillary electrometer results unless the extra mercury-"amalgam" external potential difference is included. In the same way, the results for the calomel electrode external potential difference can be obtained, either from the Volta effect results or from the above results for the mercury-solution boundary, using the potential differences of cells. Klein and Lange found the Volta effect for the calomel electrode to be 0·16 volts ; correcting from argon to moist air, this gives about 0·10 volts, and using 0·25 volts for the air-water potential difference, we get the external potential difference of the calomel electrode to be —0·15 volts.

6. *Results with Silver.*

Results with silver have already been quoted in Part V., and we shall now discuss them. A large number of various experiments have been carried out to determine the "single potential difference," or external potential difference, at a silver-solution junction. Silver is a very suitable metal to use for such experiments, since the condition of zero external potential difference at the silver electrode can be obtained merely by using the correct concentration of silver nitrate in contact with the silver.

Very frequently, a cell has been built up with the "null" silver-silver nitrate solution junction, a salt bridge and a calomel electrode, and it has been assumed that the only potential difference contributing to the total effect of the cell is that of the calomel electrode ; of course, there is also an external potential difference at the silver-mercury junction, and so the results with silver do not give a measure of the external potential difference of the calomel electrode.

The first measurements with silver appear to have been those of Billitzer (15), who used finely powdered silver in place of mercury in the dropping electrode, and he also found conditions for no movement of silver fibres or colloidal silver particles in electric fields in solutions. These results agreed in giving a potential difference of about —0·2 volts for the circuit completed with a salt bridge and a calomel electrode. Zahrzewski (16) found that a potential difference was set up by the movement of solutions in silver capillary tubes. With silver nitrate, the sign of the potential difference depends on the concentration and the sign changes at a concentration of about 1/3000 normal.

Proskurnin and Frumkin (17) brought gas-free silver suddenly into contact with silver nitrate solution, and found a current to flow ; when the concentration reaches a critical value, between 10^{-4} and 10^{-5} normal, the current changes sign. The completion of the circuit as above gives a potential difference of —0·23 volts.

Garrison (18) set up an arrangement similar to a quadrant electrometer, and made of silver, inside a solution of silver nitrate. On applying a potential difference between the quadrants, the needle moved if it carried any charge due to an external potential difference between it and the solution. Garrison found the concentration to give zero motion, and the completion of the circuit as above gave a value of between —0·1 and —0·2 volts.

Bennewitz and Schulz (19) scraped a silver electrode in silver nitrate solution with a diamond and found a current to flow, renewing the double layer at the boundary. For the critical concentration no current flows, and completion of the circuit as above gives a value of —0·20 volts; Bennewitz and Bigalke (20) repeated the method and obtained a result of —0·32 volts.

Although the above results with silver are not very consistent, we can take a value of —0·25 volts for the potential difference of a cell composed of silver, a null silver nitrate solution, concentration about 5×10^{-5} normal, a salt bridge and a calomel electrode. This had led to the belief that the single potential difference of the calomel electrode was —0·25 volts, and the fair agreement of this with the results for the mercury-mercurous ions boundary gave support to this idea. But, of course, this neglects the external potential difference at the silver-mercury junction.

A cell with silver, normal silver nitrate, a salt bridge and a calomel electrode gives a potential difference of —0·52 volts (or approximately —0·50 volts). Therefore, the cell of silver, null silver nitrate solution, salt bridge, a normal silver nitrate solution and silver gives a potential difference of —0·25 volts; thus the "single potential difference" or external potential difference at a normal silver electrode is 0·25 volts, the silver being positively charged.

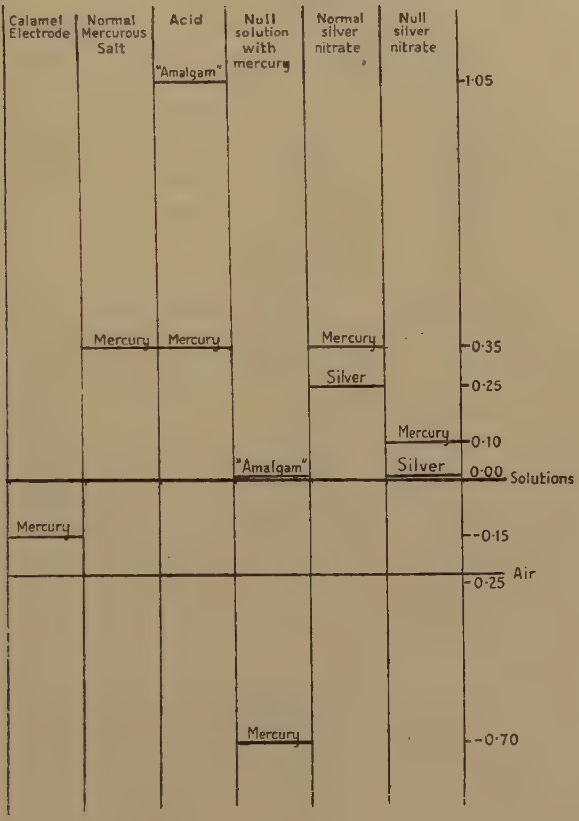
The value of —0·25 volts for the cell of the null silver electrode and the calomel electrode can give agreement with the more direct results for the calomel electrode by the use of a suitable value for the silver-mercury external potential difference. If the external potential difference of the calomel electrode is 0·56 volts, we must have a silver-mercury external potential difference of about —0·80 volts (mercury being positive). But if the potential difference of the calomel electrode is about —0·15 volts, the silver-mercury potential difference need only be about —0·10 volts.

The metal-metal external potential difference can be obtained, as discussed in Part II., from the photoelectric work functions; Klein and Lange (13) have used experimental values and give a result of —0·16 volts for the silver-mercury potential difference. This is in much closer agreement with the value of —0·10 volts than the value of —0·80 volts, and so confirms the idea that the true external potential difference of the calomel electrode is about —0·15 volts.

A consistent scheme can be built up as shown in the figure, and is probably accurate to about 0·05 volts. The zero line is taken as that for solutions, which are all connected by salt bridges, giving no external potential difference between them.

The potential difference of a cell can be found by comparing the levels of the same metal with two different solutions, *e. g.* the cell of silver-normal silver nitrate and a calomel electrode gives a potential difference of about 0·50 volts, including the effect of the mercury-silver junction.

The Volta effect of a system can be found from the difference of level of a metal and air. It can be seen that the neglect of the water-air external potential difference would give results for the Volta effect in poor accord with experiment; for example, with silver and normal silver nitrate the Volta effect would be 0·25 volts, instead of 0·50 volts, to compare with the experimental value of 0·53 volts, perhaps to be reduced by the change from argon to moist air.


Calomel Electrode
Normal Mercurous Salt
Acid
Null solution with mercury
Normal silver nitrate
Null silver nitrate
"Amalgam"
1·05
Mercury
Mercury
Mercury
0·35
Silver
0·25
Mercury
0·10
"Amalgam"
Silver
0·00 Solutions
Mercury
-0·15
Air
-0·25
Mercury
-0·70

If the idea of the " amalgam " is not accepted, then the levels of mercury in contact with the calomel electrode and with acid are to be raised by 0·70 volts, requiring Volta effects of 0·80 volts and 1·30 volts, or, neglecting the water-air potential difference, 0·55 and 1·05 volts, whereas Klein and Lange (13) found 0·16 and 0·69 volts, reduced by about 0·08 volts for moist air.

It can be seen that the scheme presented in the figure can bring about agreement between measurements of "single potential differences," and of the metal-electrolyte Volta effect, only if we accept the results for the water-air potential difference and the "amalgam"-mercury potential difference.

7. *Electrodes of Other Metals.*

The methods used for mercury are available only when the metal used is liquid. Apart from mercury, the only metal which is liquid at convenient temperatures is gallium, which has been used by Frumkin and Gorodatzkaja (21) and by Murtazajew and Gorodatzkaja (22), who found a maximum surface tension at a gallium-electrolyte junction when the circuit completed with a calomel electrode gave a potential difference of 0·9 volts. To fit with the data used in the figure, we should expect a gallium-mercury potential difference of about 0·75 volts ; the only results are for solid gallium, which would give a smaller value.

Amalgams have been used by Rothmund (23), who found that electrocapillary maxima gave agreement with potential differences of cells, but Meyer (24) found need to consider metal-metal potential differences. Frumkin and Gorodatzkaja (25) used thallium amalgam and found the potential for maximum surface tension to vary with the amount of thallium, reaching a limiting value which gave a potential difference with a calomel electrode of 0·93 volts ; if there is no "amalgam," as with pure mercury, this would mean a thallium amalgam-mercury potential difference of about 0·78 volts ; photoelectric results for pure solid thallium give 0·84 volts. Poincaré (26) found that tin amalgam in a capillary electrometer gave a maximum with zero applied potential difference, and Christiansen (7) found drops of zinc amalgam in zinc sulphate solution to be uncharged.

Billitzer (15) used platinum and gold as well as silver and found similar results. Garrison (18) used copper in place of silver and found it impossible to obtain a solution sufficiently concentrated to give a null solution, so that a copper/copper sulphate junction always has the copper negative. Bennewitz and Schultz (19), by indirect methods, obtained a value of —0·19 volts for the cell with a null copper electrode and a calomel electrode. A cell composed of copper, normal copper sulphate and a calomel electrode has a potential difference of —0·05 volts. These results suggest that the copper/normal copper sulphate external potential difference is about —0·15 volts and the mercury/copper potential difference +0·05 volts ; this last is not in good agreement with the results chosen by Klein and Lange (13), but does agree with the general run of Volta effect results for copper and silver. One would expect a Volta effect for copper and normal copper sulphate of about 0·10 volts ; results include —0·06

(Guglielmo (27)), $-0{\cdot}02$ (for $\frac{\text{N}}{2}$ solution, Bancroft and Porter (28)), $+0{\cdot}31$ (Moffitt (29)) and $+0{\cdot}18$ (Andauer and Lange (30)).

Klein and Lange (13) have used results from photoelectric data to give metal-metal potential differences and results from the potential differences of cells, and have deduced the metal-electrolyte potential differences; as we have seen in the cases of silver and mercury, a correction is needed for the air-water interface and this amounts to 0·25 volts, perhaps more since there is the change from argon to moist air. Making this correction we find that the zinc/zinc sulphate external potential difference amounts to about $-0{\cdot}20$, agreeing with the idea that zinc is always negative to the solution. An interesting metal to consider is cadmium, which, if the photoelectric data are correct, would give a null solution for some concentration of cadmium ions greater than $\frac{\text{N}}{100}$; methods like those of Garrison or Bennewitz and Schulz should be available and might be used to verify the conclusions arrived at in the present work.

8. *Conclusion.*

The so-called " single potential differences " have been discussed and a consistent interpretation of the results has been given, assuming an extra boundary at a mercury-acid surface and the electrolyte-air potential difference of about 0·25 volts. Results for other metals are discussed.

References.

(1) Lippmann, *Ann. der Phys.* v. p. 494 (1875).
(2) Kucera, *Ann. der Phys.* xi. pp. 529, 698 (1903).
(3) Craxford, Phil. Mag. xvi. p. 66 (1933).
(4) Palmaer, *Zeit. Phys. Chem.* xxv. p. 265 ; xxviii. p. 259 ; xxxvi. p. 664 (1898–1901).
(5) Palmaer, *Zeit. Elektrochem.* ix. p. 754 (1903) ; *Zeit. Phys. Chem.* lix. p. 129 (1907).
(6) Heyrovsky, Proc. Roy. Soc. cii. p. 628 (1923).
(7) Christiansen, *Ann. der Phys.* xii. p. 1072 (1903).
(8) Frumkin, J. Russ. Phys. Chem. Soc. xlix. p. 207 (1917).
(9) Bordforss, *Zeit. Elektrochem.* xxix. p. 121 (1923).
(10) Craxford, Phil. Mag. xvi. p. 268 (1933).
(11) Bennewitz and Delijannis, *Zeit. Phys. Chem.* cxxv. p. 144 (1927).
(12) Bennewitz and Küchler, *Zeit. Phys. Chem.* cliii. p. 443 (1931).
(13) Klein and Lange, *Zeit. Elektrochem.* xliii. p. 570 (1937).
(14) Chalmers and Pasquill, Phil. Mag. xxiii. p. 88 (1937).
(15) Billitzer, *Zeit. Elektrochem.* viii. p. 638 ; *Ann. der Phys.* xi. pp. 902, 937 ; xiii. p. 827 ; *Zeit. Phys. Chem.* xlviii. pp. 513, 542 ; li. p. 166 (1902–5).
(16) Zahrzewski, *Phys. Zeit.* ii. p. 146 (1902).
(17) Proskurnin and Frumkin, *Zeit. Phys. Chem.* clv. p. 29 (1931).
(18) Garrison, J. Amer. Chem. Soc. xlv. p. 37 (1923).
(19) Bennewitz and Schulz, *Zeit. Phys. Chem.* cxxiv. p. 115 (1926).
(20) Bennewitz and Bigalke, *Zeit. Phys. Chem.* cliv. p. 113 (1931).
(21) Frumkin and Gorodatzkaja, *Zeit. Phys. Chem.* cxxxvi. p. 217 (1928).
(22) Murtazajew and Gorodatzkaja, *Acta Physiochemica*, iv. p. 75 (1936).

(23) Rothmund, *Zeit. Phys. Chem.* xv. p. 1 (1894).
(24) Meyer, *Wied. Ann.* lvi. p. 680 (1895).
(25) Frumkin and Gorodatzkaja, *Zeit. Phys. Chem.* cxxxvi. p. 451 (1928).
(26) Poincaré, *Comp. Rend.* cxxxiv. p. 527 (1902).
(27) Guglielmo, *Accad. Lincei Atti.* xviii. p. 350 (1909).
(28) Bancroft and Porter, J. Phys. Chem. xl. p. 1201 (1936).
(29) Moffitt, Phys. Rev. ii. p. 95 (1913).
(30) Andauer and Lange, *Zeit. Phys. Chem.* clxvi. p. 219 (1933).

LXIV. *Contact Potentials.*—Part VII. *Chemical and Thermal Effects.*

By J. Alan Chalmers, M.A., Ph.D., Lecturer in Physics in the Durham Colleges of the University of Durham *.

[Received March 20, 1942.]

1. *Introduction.*

In previous parts the problems of contact potentials have been discussed by reference to changes in the thermodynamic potential. But, as has been discussed in Parts III. and IV., the origin of the potential differences may be connected partly with chemical changes and partly with thermal changes.

It is the purpose of the present paper to discuss the phenomena of contact potentials in a slightly different way from that of the previous parts in order to bring out the distinction between the chemical and thermal origins of the potential differences.

2. *The Thermodynamic Potential.*

The thermodynamic potential, G, of a system is defined by $G=H-TS$, where H is the quantity known as the "total heat" or "enthalpy," T is the temperature, and S the entropy. For a change at constant temperature, $\Delta G=\Delta H-T\Delta S=\Delta H-\Delta Q$ where ΔQ is the change in heat content, or the energy that must be provided from outside in the form of heat in order to keep the temperature constant. ΔG represents the energy change in the system when the temperature remains constant, while ΔH is that when the change is adiabatic, the pressure remaining constant in all cases.

Since $S=-\left(\frac{\partial G}{\partial T}\right)_p$, we get the Thomson equation :—

$$\Delta G=\Delta H+T\left(\frac{\partial \Delta G}{\partial T}\right)_p .$$

This, when applied to the Voltaic cell, is known as the Gibbs-Helmholtz equation, and gives a relation between ΔQ, ΔG and ΔH.

---

* Communicated by the Author.

The division of a thermodynamic potential difference ΔG into ΔH, which can be termed a "chemical" part and ΔQ, a thermal part, can be most clearly brought out if contact potentials are discussed by reference to changes in H, rather than changes in G.

3. *The Enthalpy.*

The quantity, H, defined as $E+pv$, where E is the internal energy, p the pressure and v the volume, is most commonly known as the "total heat." From the point of view of the present discussion, in which the changes in H are sharply distinguished from heat changes, the term "total heat" is singularly unfortunate; the only other term used is "enthalpy," and, in spite of its awkwardness, this is the term that we shall adopt.

We can discuss the phenomena of contact potentials in terms of enthalpy differences instead of potential differences, and enthalpy differences can be defined and named in just the same way as potential differences; for example, there is a removal enthalpy difference which is the change of enthalpy in removing a particle from a phase. For notation, we shall use the same letters for enthalpy differences as for potential differences but we shall use capital *italics* for enthalpy differences, *e. g.*, A_{12} is the external enthalpy difference and ${}_2C_1$ is a change of carrier enthalpy difference.

It is easy to see that, just as for the potential differences, so for the enthalpy differences, the total for a complete system can be built up from external enthalpy differences or from internal enthalpy differences. We shall see later that external enthalpy differences and external potential differences are the same, and so the total of the external enthalpy differences of a system is equal to the measured potential difference.

4. *Equilibrium at a Boundary.*

At a boundary at which there is equilibrium, there will be no change in thermodynamic potential when a charge passes across, *i. e.*, $\Delta G=0$.

Thus $\Delta H-\Delta Q=0$, *i. e.*, the enthalpy change is equal to the heat change, or all the energy supplied is used up as heat.

A little care is needed in regard to the signs of the quantities concerned. If ΔG is positive there is an increase in thermodynamic potential and energy must be supplied from outside to make the change isothermally. If ΔH is positive, there is an increase in enthalpy and a supply of energy from outside for an adiabatic change. If ΔH is greater than ΔG, the energy supplied for the adiabatic change is more than sufficient to make the change isothermally, and so there is a surplus of energy which appears as heat, ΔQ, which is thus positive. When ΔG is zero, a positive ΔH gives a positive ΔQ, *i. e.*, a heating effect. When heat is supplied from outside to keep the temperature constant, this must mean that an adiabatic change will cause cooling, *i. e.*, ΔG is greater than ΔH, so that ΔQ is negative.

At a boundary at which there is equilibrium, there will be an enthalpy difference in crossing the boundary, which is a transfer enthalpy difference M_{12}, and is equal to the heat produced on crossing the boundary. If conditions of the passage of a virtual current were the same as those of a continuous current, then M_{12} would be the Peltier coefficient of the junction. But, as has been pointed out by the author (1), the thermodynamic calculation of the Peltier effect, a calculation of what is here called M_{12}, does not give the same result as the "transport" calculation by Wilson (2).

Kelvin's (3) thermodynamic argument, which amounts to the use of $\Delta \mathrm{Q}=-\mathrm{T}\frac{\partial \Delta \mathrm{G}}{\partial \mathrm{T}}$, shows that

$$M_{12}=\mathrm{T}\frac{\partial}{\partial \mathrm{T}}\mathrm{A}_{12}.$$

Now, in a similar way,

$${}_2D_1={}_2\mathrm{D}_1-\mathrm{T}\frac{\partial}{\partial \mathrm{T}}{}_2\mathrm{D}_1,$$

and

$${}_1D_2={}_1\mathrm{D}_2-\mathrm{T}\frac{\partial}{\partial \mathrm{T}}{}_1\mathrm{D}_2.$$

From the definitions,

$$A_{12}={}_1D_2-{}_2D_1+M_{12},$$

and

$$\mathrm{A}_{12}={}_1\mathrm{D}_2-{}_2\mathrm{D}_1+\mathrm{M}_{12},$$

and M_{12} is zero.

So we see that :—

$$\begin{aligned}\mathrm{A}_{12}-A_{12}&=\mathrm{T}\frac{\partial}{\partial \mathrm{T}}({}_1\mathrm{D}_2-{}_2\mathrm{D}_1)-M_{12},\\&=\mathrm{T}\frac{\partial}{\partial \mathrm{T}}(\mathrm{A}_{12})-M_{12},\\&=0.\end{aligned}$$

Thus we see that the external enthalpy difference is equal to the external potential difference at a boundary where there is equilibrium.

It can easily be seen that the same is not true for the internal enthalpy and potential differences, but that : $\mathrm{B}_{12}-B_{12}=\mathrm{T}\frac{\partial}{\partial \mathrm{T}}(\mathrm{D}_1-\mathrm{D}_2)$.

Some of the relations expressed in this section are illustrated in the figure.

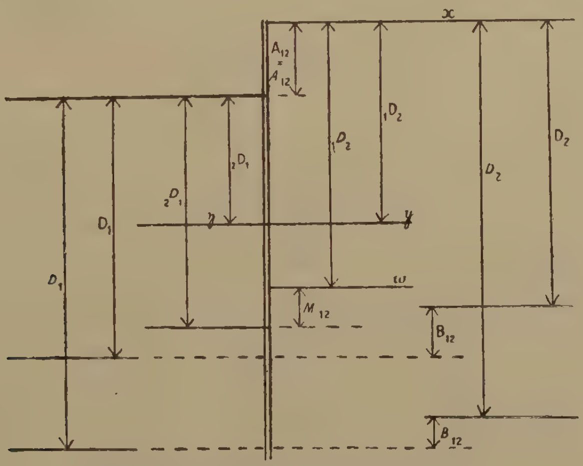
5. *Dielectric Boundaries.*

At a dielectric boundary, the external potential difference is zero, and the same must apply to the external enthalpy difference. Since there is not equilibrium, we can no longer say that the transfer enthalpy difference, M_{13}, is equal to a heating effect, but M_{13} $(={}_3D_1)$ is an enthalpy difference which is different from the heat change, ${}_3D_1-{}_3\mathrm{D}_1$.

If we have a system of two conductors and a dielectric, *i. e.* the Volta effect, then, by considering enthalpy differences round the circuit, we see that the only one which gives a heating effect is that at the boundary of the two conductors ; the Peltier effect is, of course, equivalent to this but for the distinction between " virtual " and continuous current.

6. *The Total Enthalpy Difference of a System.*

We have seen that the external enthalpy difference at any boundary is the same as the external potential difference, and so the total enthalpy difference of a system, which is the sum of the external enthalpy differences, is equal the total potential difference, which is the sum of the external potential differences. This result can be seen from another method of argument distinct from that which led to $A_{12}=\mathrm{A}_{12}$. If we consider the passage of a charge across a vacuum, it is quite immaterial to the work done whether it is an isothermal or an adiabatic process, and thus the differences of potential and enthalpy are the same. The passage across a vacuum can be considered to refer both to the external potential difference and also to the total potential difference which can be considered to be measured by the potential difference across a vacuum break in one conductor.



Enthalpy and potential differences.

Since the Gibbs-Helmholtz equation certainly holds for a complete system (it was enunciated for the chemical cell), we have an apparent anomaly, since

$$\Sigma\Delta\mathrm{G}=\Sigma\Delta\mathrm{H},$$

and so we should expect $\Sigma\Delta\mathrm{Q}$ to be zero, whereas, in fact, we know that there are cases, *e. g.* the concentration cell, in which the whole of the potential difference is supplied by thermal energy.

The answer to this difficulty must lie in the fact that what we have called the external enthalpy difference, A_{12}, is comprised not only of the two enthalpy differences, D_2 and ${}_2D_1$, but also of the transfer enthalpy difference M_{12}, which is a thermal effect and therefore cannot truly be described as an enthalpy difference contributing to $\Sigma\Delta H$. So that, from the point of view of the Gibbs-Helmholtz equation,

$$\begin{aligned}\Sigma\Delta G=\Sigma A_{12}&=\Sigma(A_{12}-M_{12})+\Sigma M_{12}\\ &=\Sigma\Delta H \qquad\quad +\Sigma\Delta Q.\end{aligned}$$

It is necessary to enquire whether there is any similar thermal effect masquerading as an enthalpy difference at a dielectric boundary. We can see that at a dielectric boundary there is a removal of a charge to be considered, but that there is an equal and opposite effect due to the arrival of this same charge in the same conductor, in the form of the effect of change of carrier ; so any thermal effects which may exist will, in fact, cancel out in considering the Gibbs-Helmholtz equation for the whole circuit, and this has been tacitly assumed in deriving the result :

$$M_{12}=T\frac{\partial}{\partial T}A_{12}.$$

The sum of the internal enthalpy differences of a system must also equal the total potential difference of the system, but internal enthalpy differences are not the same as internal potential differences, though the sums for a complete circuit must be the same.

Thus the total potential difference of a system can be split up in three ways :—(1) as external potential or enthalpy differences, (2) as internal potential differences, and (3) as internal enthalpy differences.

In the case of the Volta effect of two metals, the external potential or enthalpy differences give an effect at the metal-metal junction but none at the dielectric boundaries, the internal potential differences give no effect at the metal-metal junction, the whole occurring at the dielectric boundaries, while the internal enthalpy difference gives most of the effect at the dielectric boundaries, a small part, M_{12}, corresponding nearly to the Peltier effect, remaining at the metal-metal junction.

7. *Chemical Changes.*

It is in the Voltaic cell that the distinction between chemical and thermal effects is most important and most noticeable.

If the ions in a solution behave like particles in a perfect gas, then the change of enthalpy in removing any ion does not depend on the concentration, though the change of potential does ; in a solution, the value of H per particle is independent of concentration, but the value of S, and hence of G, per particle does depend on the concentration.

Thus, if we concern ourselves only with enthalpy changes, we shall get no effects of concentration, and, since $\Sigma\Delta H$ represents the " chemical " part of the energy supplied, we see that, as might be expected, there is no effect of concentration on this. The effect of concentration must appear entirely in the " thermal " part of the energy.

Using the Gibbs-Helmholtz equation :

$$\varDelta G=\varDelta H-\varDelta Q=\varDelta H+T\frac{\partial \varDelta G}{\partial T},$$

we see that any term proportional to T in $\varDelta H$ requires a term proportional to T log T in $\varDelta Q$, while there can be terms proportional to T in $\varDelta Q$ with no corresponding term in $\varDelta H$. The concentration only appears in $\varDelta G$ in a term proportional to T log c, so that this must form part of $\varDelta Q$, not of $\varDelta H$.

In the concentration cell $\Sigma\varDelta H$ is zero, and so the sum of the enthalpy changes must be zero, and the whole of the potential difference is due to thermal effects and is therefore ΣM_{12}.

At the metal-electrolyte junctions, we see that ${}_1D_3$ and ${}_1D_4$ are equal and do not depend on concentration ; clearly ${}_3D_1$ and ${}_4D_1$ are equal, as well as ${}_3\mathrm{D}_1$ and ${}_4\mathrm{D}_1$. But ${}_1\mathrm{D}_3$ and ${}_1\mathrm{D}_4$ do depend on concentration and include terms kT log c. So we see that $M_{13}-M_{14}=k\mathrm{T}\log c_3/c_4$; this is clear from the fig. 1, for a change of concentration on the right-hand side raises the level " x," the level " y " being necessarily the same as the level " z," since M_{12} remains zero. Thus the level " w " is raised by the same amount as " x," since ${}_1D_2$ is unaltered.

Considering now the junction of the two electrolytes in the concentration cell, we see that, since ${}_4D_3={}_3D_4$, $M_{12}=A_{12}=\mathrm{A}_{12}$; and, since there is no change of carrier, $B_{12}=M_{12}$. Thus, at the junction of two solutions of the same electrolyte with different concentrations, the " contact potential difference " can be considered to have a definite value, as measured by the external potential difference, the external enthalpy difference, the internal enthalpy difference or the transfer enthalpy difference, which gives the heating effect ; or it can be considered to be zero, as measured by the internal potential difference or by the transfer potential difference.

8. *Conclusion.*

The chemical and thermal origins of the potential differences of contact potential systems are best distinguished by the consideration of enthalpy differences rather than potential differences. This gives other ways of dividing up the total potential difference of a system, namely, according to external enthalpy differences or according to internal enthalpy differences. External enthalpy differences are the same as external potential differences, but internal enthalpy differences differ from internal potential differences. The effects of concentration all occur as " thermal " rather than " chemical " effects.

References.

(1) Chalmers, Proc. Camb. Phil. Soc. xxxiii. p. 521 (1939).
(2) Wilson, ' The Theory of Metals,' p. 177 (1936).
(3) Kelvin, Math. and Phys. Papers, v. p. 29.

LXV. *The Application of the Laplace Transformation to a Problem on Elastic Vibrations.*

By C. J. Tranter, M.A.*

[Received April 16, 1942.]

1. Recent papers by Carslaw, Jaeger and others† have shown how effectively the Laplace Transformation will handle boundary value problems, many of their illustrations being taken from problems in the conduction of heat. In the present paper, the vibrations in a thick elastic cylinder under an internal pressure varying with the time are considered. The method developed by Carslaw and Jaeger is particularly suitable for this problem, which is of some practical importance. The advantage of the method is that the amplitudes of the vibrations are given easily, in contrast to the classical method which requires some theorem of expansion in normal functions analogous to Fourier's theorem. The solution to the similar problem for the sphere is also given.

2. *Forced vibrations in a thick cylinder.*—We consider a thick cylinder, supposed isotropic and homogeneous, of infinite length and of radii r_0, r_1. The cylinder is under internal pressure $A(1-\cos \omega t)$, A, ω, being constants and t the time, and the external pressure is assumed to be zero.

If U denotes the radial displacement at radius r and time t

$$\frac{\partial^2 U}{\partial r^2}+\frac{1}{r}\frac{\partial U}{\partial r}-\frac{U}{r^2}=\frac{1}{a^2}\frac{\partial^2 U}{\partial t^2}, \qquad r_0<r<r_1,\ t>0, \quad \ldots \quad (1)$$

where

$$a^2=(\lambda+2\mu)/\rho, \quad \ldots \quad (2)$$

ρ is the density of the cylinder and λ, μ are given in terms of Young's Modulus (E) and Poisson's Ratio (σ) by

$$\lambda=\frac{E\sigma}{(1+\sigma)(1-2\sigma)},\ \mu=\frac{E}{2(1+\sigma)}. \quad \ldots \quad (3)$$

The boundary conditions are

$$\left.\begin{aligned}(\lambda+2\mu)\frac{\partial U}{\partial r}+\lambda\frac{U}{r}&\to -A(1-\cos\omega t), \text{ when } r\to r_0,\ t>0\\ (\lambda+2\mu)\frac{\partial U}{\partial r}+\lambda\frac{U}{r}&\to 0, \qquad\qquad \text{ when } r\to r_1,\ t>0\end{aligned}\right\}, \quad . \quad (4)$$

---

* Communicated by the Author.

† See, for instance, H. S. Carslaw and J. C. Jaeger, "Some Two-Dimensional Problems in Conduction of Heat with Circular Symmetry," Proc. London Math. Soc., ser. 2, vol. xlvi. (1940).

the initial conditions being

$$\left.\begin{array}{l} \mathrm{U}\to 0 \\ \dfrac{\partial \mathrm{U}}{\partial t}\to 0 \end{array}\right\} \text{when } t\to 0,\ r_0<r<r_1. \quad \ldots\ldots \quad (5)$$

Multiplying equation (1) by e^{-pt}, where $p>0$, integrating with respect to t from 0 to ∞ and writing

$$\overline{\mathrm{U}}=\int_0^\infty e^{-pt}\mathrm{U}dt, \quad \ldots\ldots\ldots \quad (6)$$

we have *

$$\frac{d^2\overline{\mathrm{U}}}{dr^2}+\frac{1}{r}\frac{d\overline{\mathrm{U}}}{dr}-\left(\frac{1}{r^2}+\frac{p^2}{a^2}\right)\overline{\mathrm{U}}=0,\ r_0<r<r_1, \quad \ldots\ldots \quad (7)$$

since

$$\int_0^\infty e^{-pt}\frac{\partial^2\mathrm{U}}{\partial t^2}dt=\left[e^{-pt}\left(\frac{\partial \mathrm{U}}{\partial t}+p\mathrm{U}\right)\right]_0^\infty+p^2\int_0^\infty e^{-pt}\mathrm{U}dt,$$

and the first term on the right-hand side vanishes if we assume that the limits as t tends to infinity of $e^{-pt}\mathrm{U}$ and $e^{-pt}(\partial\mathrm{U}/\partial t)$ are both zero, conditions (5) showing that the term vanishes at the lower limit.

Since

$$\int_0^\infty e^{-pt}(1-\cos\omega t)dt=\frac{\omega^2}{p(p^2+\omega^2)}$$

the boundary conditions (4) become

$$\left.\begin{array}{ll} (\lambda+2\mu)\dfrac{d\overline{\mathrm{U}}}{dr}+\lambda\dfrac{\overline{\mathrm{U}}}{r}=-\dfrac{\mathrm{A}\omega^2}{p(p^2+\omega^2)}, & \text{when } r=r_0, \\ (\lambda+2\mu)\dfrac{d\overline{\mathrm{U}}}{dr}+\lambda\dfrac{\overline{\mathrm{U}}}{r}=0, & \text{when } r=r_1 \end{array}\right\} \quad \ldots \quad (8)$$

The solution of (7) is

$$\overline{\mathrm{U}}=\mathrm{CI}_1(pr/a)+\mathrm{DK}_1(pr/a). \quad \ldots\ldots\ldots \quad (9)$$

Using the results †

$$\left.\begin{array}{l} z\mathrm{I}_1{}'(z)+\mathrm{I}_1(z)=z\mathrm{I}_0(z) \\ z\mathrm{K}_1{}'(z)+\mathrm{K}_1(z)=-z\mathrm{K}_0(z) \end{array}\right\} \quad \ldots\ldots \quad (10)$$

and substituting in (8) we have

$$\mathrm{C}\left[\mathrm{I}_1\left(\frac{pr_0}{a}\right)-\gamma\frac{pr_0}{a}\mathrm{I}_0\left(\frac{pr_0}{a}\right)\right]+\mathrm{D}\left[\mathrm{K}_1\left(\frac{pr_0}{a}\right)+\gamma\frac{pr_0}{a}\mathrm{K}_0\left(\frac{pr_0}{a}\right)\right]=\frac{\mathrm{A}\omega^2r_0}{2\mu p(p^2+\omega^2)} \quad \ldots\ldots \quad (11)$$

$$\mathrm{C}\left[\mathrm{I}_1\left(\frac{pr_1}{a}\right)-\gamma\frac{pr_1}{a}\mathrm{I}_0\left(\frac{pr_1}{a}\right)\right]+\mathrm{D}\left[\mathrm{K}_1\left(\frac{pr_1}{a}\right)+\gamma\frac{pr_1}{a}\mathrm{K}_0\left(\frac{pr_1}{a}\right)\right]=0, \quad (12)$$

where

$$\gamma=\frac{\lambda+2\mu}{2\mu}. \quad \ldots\ldots\ldots \quad (13)$$

---

* In deriving the "subsidiary equations" (7) and (8), assumptions as to the commutability of certain limiting operations and that the order of differentiation and integration can be changed have been made. Similar assumptions are made by Carslaw and Jaeger, *loc. cit.* p. 363.

† Watson, "Theory of Bessel Functions," 1922, § 3.71.

Solving and substituting in (9)

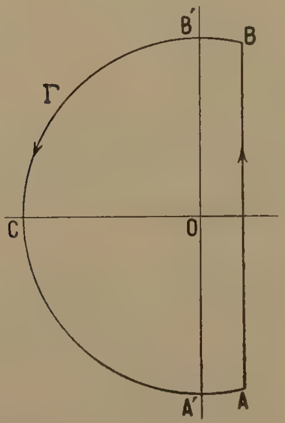
$$\overline{\mathrm{U}}=\frac{\mathrm{A}\omega^2 r_0}{2\mu p(p^2+\omega^2)}\frac{\mathrm{F}(p)}{\mathrm{G}(p)}, \qquad \text{(14)}$$

where
$$\mathrm{F}(p)=\left[\mathrm{K}_1\left(\frac{pr_1}{a}\right)+\gamma\frac{pr_1}{a}\mathrm{K}_0\left(\frac{pr_1}{a}\right)\right]\mathrm{I}_1\left(\frac{pr}{a}\right)$$
$$-\left[\mathrm{I}_1\left(\frac{pr_1}{a}\right)-\gamma\frac{pr_1}{a}\mathrm{I}_0\left(\frac{pr_1}{a}\right)\right]\mathrm{K}_1\left(\frac{pr}{a}\right), \qquad \text{(15)}$$

and
$$\mathrm{G}(p)=\left[\mathrm{I}_1\left(\frac{pr_0}{a}\right)-\gamma\frac{pr_0}{a}\mathrm{I}_0\left(\frac{pr_0}{a}\right)\right]\left[\mathrm{K}_1\left(\frac{pr_1}{a}\right)+\gamma\frac{pr_1}{a}\mathrm{K}_0\left(\frac{pr_1}{a}\right)\right]$$
$$-\left[\mathrm{K}_1\left(\frac{pr_0}{a}\right)+\gamma\frac{pr_0}{a}\mathrm{K}_0\left(\frac{pr_0}{a}\right)\right]\left[\mathrm{I}_1\left(\frac{pr_1}{a}\right)-\gamma\frac{pr_1}{a}\mathrm{I}_0\left(\frac{pr_1}{a}\right)\right]. \qquad \text{(16)}$$

Thus, using the inversion theorem *, we have

$$\mathrm{U}=\frac{\mathrm{A}\omega^2 r_0}{4\mu\pi i}\int_{c-i\infty}^{c+i\infty}\frac{e^{\zeta t}}{\zeta^2+\omega^2}\frac{\mathrm{F}(\zeta)}{\mathrm{G}(\zeta)}\frac{d\zeta}{\zeta}. \qquad \text{(17)}$$



To evaluate (17), consider the integral taken round the closed contour of the figure, consisting of the line AB at distance c from the imaginary axis and a portion of a circle Γ, centre the origin and radius $\mathrm{R}=\frac{(n+\frac{1}{2})a\pi}{r_1-r_0}$ †. It can be shown that the limit of the integral round the circle Γ tends to zero as n tends to infinity. Cauchy's theorem enables us to replace the line integral in (17) by $2\pi i$ times the sum of the residues at its poles.

---

* This states that, subject to certain restrictions on $f(p)$, if

$$f(p)=\int_0^\infty e^{-pt}\phi(t)dt, \qquad \phi(t)=(1/2\pi i)\int_{c-i\infty}^{c+i\infty}e^{\zeta t}f(\zeta)d\zeta,$$

where c is a real positive constant.

† With this choice of R, the contour avoids all poles of the integrand of (17), see § 3.

The poles are at $\zeta=0$, $\pm i\omega$ and at the roots of $\mathrm{G}(\zeta)=0$. To find these roots put $\zeta=i\alpha$, so that

$$\begin{aligned}\mathrm{G}(\zeta)=\mathrm{G}(i\alpha)=\mathrm{G}(-i\alpha)=&\\ -\frac{\pi}{2}\Big[\Big\{\mathrm{J}_1\Big(\frac{\alpha r_0}{a}\Big)-\gamma\frac{\alpha r_0}{a}\mathrm{J}_0\Big(\frac{\alpha r_0}{a}\Big)\Big\}&\Big\{\mathrm{Y}_1\Big(\frac{\alpha r_1}{a}\Big)-\gamma\frac{\alpha r_1}{a}\mathrm{Y}_0\Big(\frac{\alpha r_1}{a}\Big)\Big\}\\ -\Big\{\mathrm{Y}_1\Big(\frac{\alpha r_0}{a}\Big)-\gamma\frac{\alpha r_0}{a}\mathrm{Y}_0\Big(\frac{\alpha r_0}{a}\Big)\Big\}&\Big\{\mathrm{J}_1\Big(\frac{\alpha r_1}{a}\Big)-\gamma\frac{\alpha r_1}{a}\mathrm{J}_0\Big(\frac{\alpha r_1}{a}\Big)\Big\}\Big]. \quad . \quad (18)\end{aligned}$$

It will be shown later (§ 3) that generally the zeros of (18) are all real and simple. Let these be $\pm\alpha_1$, $\pm\alpha_2$, . . ., then the poles of the integrand of (17) are $\zeta=0$, $\pm i\omega$, $\pm i\alpha_s$, $s=1, 2, \ldots$.

The sum of the residues of the integrand of (17) at $\zeta=\pm i\alpha_s$ is

$$\frac{1}{(\omega^2-\alpha_s^2)}\left\{\frac{e^{i\alpha_s t}\mathrm{F}(i\alpha_s)}{\Big[\alpha\dfrac{d\mathrm{G}}{d\alpha}\Big]_{\alpha=\alpha_s}}+\frac{e^{-i\alpha_s t}\mathrm{F}(-i\alpha_s)}{\Big[\alpha\dfrac{d\mathrm{G}}{d\alpha}\Big]_{\alpha=-\alpha_s}}\right\}. \quad . \quad . \quad . \quad (19)$$

Differentiating (18) and using the recurrence formulæ, we have

$$\begin{aligned}&-\frac{2}{\pi}\frac{d\mathrm{G}}{d\alpha}\\ &=\frac{r_0}{a}\Big[\Big\{(1-\gamma)\mathrm{J}_0\Big(\frac{\alpha r_0}{a}\Big)+\frac{\gamma\alpha^2r_0^2-a^2}{a\alpha r_0}\mathrm{J}_1\Big(\frac{\alpha r_0}{a}\Big)\Big\}\Big\{\mathrm{Y}_1\Big(\frac{\alpha r_1}{a}\Big)-\gamma\frac{\alpha r_1}{a}\mathrm{Y}_0\Big(\frac{\alpha r_1}{a}\Big)\Big\}\\ &\qquad-\Big\{(1-\gamma)\mathrm{Y}_0\Big(\frac{\alpha r_0}{a}\Big)+\frac{\gamma\alpha^2r_0^2-a^2}{a\alpha r_0}\mathrm{Y}_1\Big(\frac{\alpha r_0}{a}\Big)\Big\}\Big\{\mathrm{J}_1\Big(\frac{\alpha r_1}{a}\Big)-\gamma\frac{\alpha r_1}{a}\mathrm{J}_0\Big(\frac{\alpha r_1}{a}\Big)\Big\}\Big]\\ &+\frac{r_1}{a}\Big[\Big\{\mathrm{J}_1\Big(\frac{\alpha r_0}{a}\Big)-\gamma\frac{\alpha r_0}{a}\mathrm{J}_0\Big(\frac{\alpha r_0}{a}\Big)\Big\}\Big\{(1-\gamma)\mathrm{Y}_0\Big(\frac{\alpha r_1}{a}\Big)+\frac{\gamma\alpha^2r_1^2-a^2}{a\alpha r_1}\mathrm{Y}_1\Big(\frac{\alpha r_1}{a}\Big)\Big\}\\ &\qquad-\Big\{\mathrm{Y}_1\Big(\frac{\alpha r_0}{a}\Big)-\gamma\frac{\alpha r_0}{a}\mathrm{Y}_0\Big(\frac{\alpha r_0}{a}\Big)\Big\}\Big\{(1-\gamma)\mathrm{J}_0\Big(\frac{\alpha r_1}{a}\Big)+\frac{\gamma\alpha^2r_1^2-a^2}{a\alpha r_1}\mathrm{J}_1\Big(\frac{\alpha r_1}{a}\Big)\Big\}\Big].\\ &\qquad\qquad\qquad\qquad\qquad\qquad . \quad . \quad . \quad . \quad . \quad (20)\end{aligned}$$

When $\alpha=\alpha_s$, a zero of (18),

$$\frac{\mathrm{Y}_1\Big(\dfrac{\alpha_s r_1}{a}\Big)-\gamma\dfrac{\alpha_s r_1}{a}\mathrm{Y}_0\Big(\dfrac{\alpha_s r_1}{a}\Big)}{\mathrm{Y}_1\Big(\dfrac{\alpha_s r_0}{a}\Big)-\gamma\dfrac{\alpha_s r_0}{a}\mathrm{Y}_0\Big(\dfrac{\alpha_s r_0}{a}\Big)}=\frac{\mathrm{J}_1\Big(\dfrac{\alpha_s r_1}{a}\Big)-\gamma\dfrac{\alpha_s r_1}{a}\mathrm{J}_0\Big(\dfrac{\alpha_s r_1}{a}\Big)}{\mathrm{J}_1\Big(\dfrac{\alpha_s r_0}{a}\Big)-\gamma\dfrac{\alpha_s r_0}{a}\mathrm{J}_0\Big(\dfrac{\alpha_s r_0}{a}\Big)}=\xi,\text{ say,} \quad (21)$$

and *

$$\mathrm{J}_0(z)\mathrm{Y}_1(z)-\mathrm{J}_1(z)\mathrm{Y}_0(z),=-\frac{2}{\pi z}, \quad . \quad . \quad . \quad . \quad . \quad (22)$$

so that, after some reduction, (20) gives

$$\Big[\alpha\frac{d\mathrm{G}}{d\alpha}\Big]_{\alpha=\alpha_s}=\xi\Big\{\frac{\gamma^2\alpha_s^2r_0^2}{a^2}-2\gamma+1\Big\}-\frac{1}{\xi}\Big\{\frac{\gamma^2\alpha_s^2r_1^2}{a^2}-2\gamma+1\Big\}=\Big[\alpha\frac{d\mathrm{G}}{d\alpha}\Big]_{\alpha=-\alpha_s}. \quad (23)$$

Again, from (15),

$$\begin{aligned}\mathrm{F}(i\alpha)=-\frac{\pi}{2}\Big[&\Big\{\mathrm{Y}_1\Big(\frac{\alpha r_1}{a}\Big)-\gamma\frac{\alpha r_1}{a}\mathrm{Y}_0\Big(\frac{\alpha r_1}{a}\Big)\Big\}\mathrm{J}_1\Big(\frac{\alpha r}{a}\Big)\\ &-\Big\{\mathrm{J}_1\Big(\frac{\alpha r_1}{a}\Big)-\gamma\frac{\alpha r_1}{a}\mathrm{J}_0\Big(\frac{\alpha r_1}{a}\Big)\Big\}\mathrm{Y}_1\Big(\frac{\alpha r}{a}\Big)\Big]=\mathrm{F}(-i\alpha), \quad (24)\end{aligned}$$

---

* Watson, *loc. cit.* § 3.63.

and, using (21), (23) and (24) in (19), the sum of the residues of the integrand of (17) at $\zeta=\pm i\alpha_s$ is

$$=\frac{2F(i\alpha_s)}{\omega^2-\alpha_s^2}\left[\frac{\left\{J_1\left(\frac{\alpha_s r_0}{a}\right)-\gamma\frac{\alpha_s r_0}{a}J_0\left(\frac{\alpha_s r_0}{a}\right)\right\}\left\{J_1\left(\frac{\alpha_s r_1}{a}\right)-\gamma\frac{\alpha_s r_1}{a}J_0\left(\frac{\alpha_s r_1}{a}\right)\right\}\cos\alpha_s t}{\left\{\frac{\gamma^2\alpha_s^2 r_0^2}{a^2}-2\gamma+1\right\}\left\{J_1\left(\frac{\alpha_s r_1}{a}\right)-\gamma\frac{\alpha_s r_1}{a}J_0\left(\frac{\alpha_s r_1}{a}\right)\right\}^2-\left\{\frac{\gamma^2\alpha_s^2 r_1^2}{a^2}-2\gamma+1\right\}\left\{J_1\left(\frac{\alpha_s r_0}{a}\right)-\gamma\frac{\alpha_s r_0}{a}J_0\left(\frac{\alpha_s r_0}{a}\right)\right\}^2}\right], \quad (25)$$

$F(i\alpha_s)$ being obtained by writing $\alpha=\alpha_s$ in (24).

The sum of the residues of the integrand of (17) at $\zeta=\pm i\omega$* is easily seen to be

$$-\frac{F(i\omega)}{\omega^2 G(i\omega)}\cos\omega t, \quad \ldots \ldots \ldots \quad (26)$$

where $F(i\omega)$, $G(i\omega)$ are obtained by writing ω for α in (18) and (24) respectively.

Expansion of the Bessel functions in (18) and (24) shows that

$$F(0)=\frac{r^2+(2\gamma-1)r_1^2}{2rr_1}, \qquad G(0)=\frac{(2\gamma-1)(r_1^2-r_0^2)}{2r_0r_1}, \quad \ldots \quad (27)$$

and the residue of the integrand of (17) at $\zeta=0$ is thus

$$\frac{1}{\omega^2}\frac{r_0}{r}\left\{\frac{r^2+(2\gamma-1)r_1^2}{(2\gamma-1)(r_1^2-r_0^2)}\right\}. \quad \ldots \ldots \ldots \quad (28)$$

Hence, finally,

$$U=\frac{Ar_0^2}{2\mu r}\left\{\frac{r^2+(2\gamma-1)r_1^2}{(2\gamma-1)(r_1^2-r_0^2)}\right\}-\frac{Ar_0}{2\mu}\frac{F(i\omega)}{G(i\omega)}\cos\omega t$$

$$+\frac{A\omega^2 r_0}{\mu}\sum_{s=1}^{\infty}\frac{F(i\alpha_s)}{\omega^2-\alpha_s^2}\left[\frac{\left\{J_1\left(\frac{\alpha_s r_0}{a}\right)-\gamma\frac{\alpha_s r_0}{a}J_0\left(\frac{\alpha_s r_0}{a}\right)\right\}\left\{J_1\left(\frac{\alpha_s r_1}{a}\right)-\gamma\frac{\alpha_s r_1}{a}J_0\left(\frac{\alpha_s r_1}{a}\right)\right\}\cos\alpha_s t}{\left\{\frac{\gamma^2\alpha_s^2 r_0^2}{a^2}-2\gamma+1\right\}\left\{J_1\left(\frac{\alpha_s r_1}{a}\right)-\gamma\frac{\alpha_s r_1}{a}J_0\left(\frac{\alpha_s r_1}{a}\right)\right\}^2-\left\{\frac{\gamma^2\alpha_s^2 r_1^2}{a^2}-2\gamma+1\right\}\left\{J_1\left(\frac{\alpha_s r_0}{a}\right)-\gamma\frac{\alpha_s r_0}{a}J_0\left(\frac{\alpha_s r_0}{a}\right)\right\}^2}\right], \quad (29)$$

where γ is given by (13), F by (24), G by (18) and the summation is over the positive zeros of (18).

The radial and hoop stresses are, with the usual notation, then given by

$$\left.\begin{aligned}\widehat{rr}&=(\lambda+2\mu)\frac{\partial U}{\partial r}+\lambda\frac{U}{r}\\ \widehat{\theta\theta}&=\lambda\frac{\partial U}{\partial r}+(\lambda+2\mu)\frac{U}{r}\end{aligned}\right\}. \quad \ldots \ldots \ldots \quad (30)$$

---

* It is assumed that ω is unequal to any α_s.

3. *The zeros of* G$(i\alpha)$.—The function F$(i\alpha)$, defined by (24), satisfies

$$\frac{\partial^2 F}{\partial r^2}+\frac{1}{r}\frac{\partial F}{\partial r}-\frac{F}{r^2}=-\frac{\alpha^2}{a^2}F. \quad \ldots \ldots \quad (31)$$

It is easy to verify that

$$(\lambda+2\mu)\frac{\partial F}{\partial r}+\lambda\frac{F}{r}=-\frac{2\mu}{r_0}G(i\alpha), \text{ when } r=r_0, \quad \ldots \quad (32)$$

$$(\lambda+2\mu)\frac{\partial F}{\partial r}+\lambda\frac{F}{r}=0, \qquad \text{when } r=r_1. \quad \ldots \quad (33)$$

If F, F$'$ correspond to α, α', equation (31) gives

$$\frac{\alpha^2-\alpha'^2}{a^2}\int_{r_0}^{r_1} rFF'dr=\left[r\left(F\frac{\partial F'}{\partial r}-F'\frac{\partial F}{\partial r}\right)\right]_{r=r_0}^{r=r_1}, \quad \ldots \quad (34)$$

and if α, α' are distinct zeros of G $(i\alpha)$, equations (32) and (33) show that

$$\int_{r_0}^{r_1} rFF'dr=0. \quad \ldots \ldots \ldots \quad (35)$$

But if α, α' are conjugate complex zeros of G$(i\alpha)$, F, F$'$ will be conjugate complex quantities and the integral in (35) cannot vanish. Thus there is no complex zero of G$(i\alpha)$.

Writing $\alpha'=\alpha+\epsilon$ in (34), and finding the limit as ϵ tends to zero, we have

$$-\frac{2\alpha}{a^2}\int_{r_0}^{r_1} rF^2dr=\left[r\left(F\frac{\partial^2 F}{\partial r\partial\alpha}-\frac{\partial F}{\partial r}\frac{\partial F}{\partial\alpha}\right)\right]_{r=r_0}^{r=r_1}. \quad \ldots \quad (36)$$

For a repeated zero, equations (32), (33) give

$$(\lambda+2\mu)\frac{\partial^2 F}{\partial r\partial\alpha}+\frac{\lambda}{r}\frac{\partial F}{\partial\alpha}=0, \text{ when } r=r_0,\ r_1, \quad \ldots \quad (37)$$

since in this case, $\partial G/\partial\alpha=0$. The right-hand side of (36) now vanishes, and since (27) shows that $\alpha=0$ is not a zero of G$(i\alpha)$, the integral on the left-hand side must vanish. But the integrand is essentially positive and the integral cannot be zero and there is thus no repeated zero of G$(i\alpha)$.

To discuss possible imaginary zeros, write $\alpha=iax/r_0$, $k=r_1/r_0$ and we find

$$G=\{K_1(x)+\gamma xK_0(x)\}\{K_1kx)+\gamma kxK_0(kx)\}$$
$$\left[\frac{I_1(x)-\gamma xI_0(x)}{K_1(x)+\gamma xK_0(x)}-\frac{I_1(kx)-\gamma kxI_0(kx)}{K_1(kx)+\gamma kxK_0(kx)}\right]. \quad (38)$$

G is an even function of x, so that it is sufficient to consider $x>0$. The asymptotic expansions of the Bessel functions show that there is no infinite zero. The first two factors of the right-hand side of (38) are essentially positive, and possible imaginary zeros of G$(i\alpha)$ are given by values of x which make the third factor of the right-hand side of (38) vanish.

Writing
$$f(x)=\frac{I_1(x)-\gamma xI_0(x)}{K_1(x)+\gamma xK_0(x)}, \quad \ldots \ldots \quad (39)$$

using the recurrence formulæ and *

$$I_0(x)K_1(x)+I_1(x)K_0(x)=\frac{1}{x},$$

we find, after some reduction,

$$f'(x)=-\frac{\gamma^2x^2+2\gamma-1}{x\{K_1(x)+\gamma xK_0(x)\}^2}. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (40)$$

Hence, if $\gamma>\frac{1}{2}$, $f'(x)$ is essentially negative for positive x, and (38) shows that G cannot vanish. For isotropic materials in practice γ is certainly greater than $\frac{1}{2}$ and the zeros of $G(i\alpha)$ are thus all real and simple.

Finally, using the asymptotic expansions in (18) we find that, for large α

$$G(i\alpha)\sim-\gamma^2\frac{\alpha}{a}\sqrt{r_1r_0}\sin\left\{(r_1-r_0)\frac{\alpha}{a}\right\},$$

and the circle Γ of the figure thus avoids all poles of the integrand of (17) as n tends to infinity.

4. *Forced vibrations in a thick sphere.*—With a similar notation we now have

$$\frac{\partial^2U}{\partial r^2}+\frac{2}{r}\frac{\partial U}{\partial r}-2\frac{U}{r^2}=\frac{1}{a^2}\frac{\partial^2U}{\partial t^2}, \qquad r_0<r<r_1,\ t>0 \quad . \quad . \qquad (41)$$

with the boundary and initial conditions

$$\left.\begin{aligned}(\lambda+2\mu)\frac{\partial U}{\partial r}+2\lambda\frac{U}{r}&\to-A(1-\cos\omega t), \text{ when } r\to r_0,\ t>0\\(\lambda+2\mu)\frac{\partial U}{\partial r}+2\lambda\frac{U}{r}&\to0, \qquad\qquad\qquad \text{when } r\to r_1,\ t>0\end{aligned}\right\}, \qquad (42)$$

$$U\to0 \text{ and } \partial U/\partial t\to0, \text{ when } t\to0,\ r_0<r<r_1. \quad . \quad . \quad . \qquad (43)$$

Proceeding as before, the " subsidiary equations " are

$$\frac{d^2\overline{U}}{dr^2}+\frac{2}{r}\frac{d\overline{U}}{dr}-\left(\frac{2}{r^2}+\frac{p^2}{a^2}\right)\overline{U}=0,\ r_0<r<r_1, \quad . \quad . \quad . \qquad (44)$$

$$\left.\begin{aligned}(\lambda+2\mu)\frac{d\overline{U}}{dr}+2\lambda\frac{\overline{U}}{r}&=-\frac{A\omega^2}{p(p^2+\omega^2)}, \text{ when } r=r_0,\\(\lambda+2\mu)\frac{d\overline{U}}{dr}+2\lambda\frac{\overline{U}}{r}&=0, \qquad\qquad\quad \text{when } r=r_1\end{aligned}\right\} \quad . \quad . \quad . \qquad (45)$$

The primitive of (44) is

$$\overline{U}=\frac{C(1-pr/a)e^{p\frac{r}{a}}+D(1+pr/a)e^{-p\frac{r}{a}}}{(pr/a)^2}. \quad . \quad . \quad . \quad . \quad . \qquad (46)$$

Substituting in (45) and now writing

$$\gamma=\frac{\lambda+2\mu}{4\mu}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (47)$$

---

* Watson, *loc. cit.* § 3.71.

we find

$$\mathrm{U}=\frac{\mathrm{A}\omega^2 r_0^3}{4\mu r^2 p(p^2+\omega^2)}\frac{\mathrm{F}(p)}{\mathrm{G}(p)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (48)$$

$\mathrm{F}(p)$, $\mathrm{G}(p)$ being now defined by

$$\mathrm{F}(p)=\left(1-\frac{pr}{a}\right)\left(\gamma\frac{p^2r_1^2}{a^2}+\frac{pr_1}{a}+1\right)e^{-p\frac{(r_1-r)}{a}}$$
$$-\left(1+\frac{pr}{a}\right)\left(\gamma\frac{p^2r_1^2}{a^2}-\frac{pr_1}{a}+1\right)e^{p\frac{(r_1-r)}{a}}, \quad (49)$$

and

$$\mathrm{G}(p)=\left\{\gamma\frac{p^2r_0^2}{a^2}-\frac{pr_0}{a}+1\right\}\left\{\gamma\frac{p^2r_1^2}{a^2}+\frac{pr_1}{a}+1\right\}e^{-p\frac{(r_1-r_0)}{a}}$$
$$-\left\{\gamma\frac{p^2r_0^2}{a^2}+\frac{pr_0}{a}+1\right\}\left\{\gamma\frac{p^2r_1^2}{a^2}-\frac{pr_1}{a}+1\right\}e^{p\frac{(r_1-r_0)}{a}}. \quad (50)$$

The inversion theorem gives

$$\mathrm{U}=\frac{\mathrm{A}\omega^2 r_0^3}{8\mu r^2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{e^{\zeta t}}{\zeta^2+\omega^2}\frac{\mathrm{F}(\zeta)}{\mathrm{G}(\zeta)}\frac{d\zeta}{\zeta}. \quad . \quad . \quad . \quad . \quad (51)$$

We evaluate the line integral as before and find, if we write $\zeta=i\alpha$,

$$\mathrm{G}(i\alpha)$$
$$=2i\left[\left\{\left(1-\gamma\frac{\alpha^2r_0^2}{a^2}\right)\sin\frac{\alpha r_0}{a}-\frac{\alpha r_0}{a}\cos\frac{\alpha r_0}{a}\right\}\left\{\left(1-\gamma\frac{\alpha^2r_1^2}{a^2}\right)\cos\frac{\alpha r_1}{a}+\frac{\alpha r_1}{a}\sin\frac{\alpha r_1}{a}\right\}\right.$$
$$\left.-\left\{\left(1-\gamma\frac{\alpha^2r_0^2}{a^2}\right)\cos\frac{\alpha r_0}{a}+\frac{\alpha r_0}{a}\sin\frac{\alpha r_0}{a}\right\}\left\{\left(1-\gamma\frac{\alpha^2r_1^2}{a^2}\right)\sin\frac{\alpha r_1}{a}-\frac{\alpha r_1}{a}\cos\frac{\alpha r_1}{a}\right\}\right]$$
$$=-\mathrm{G}(-i\alpha). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

The zeros of $\mathrm{G}(i\alpha)$, (except $\alpha=0$), can be shown by a method similar to that in § 3 to be real and simple provided $\gamma>\frac{1}{3}$, a condition certainly realized in practice. If these zeros are $\pm\alpha_s$, $s=1, 2, \ldots$ we find

$$\left[\alpha\frac{d\mathrm{G}}{d\alpha}\right]_{\alpha=\alpha_s}=2i\frac{\alpha_s^3}{a^3}\left[r_0^3\left\{\gamma^2\frac{\alpha_s^2r_0^2}{a^2}-3\gamma+1\right\}\xi-r_1^3\left\{\gamma^2\frac{\alpha_s^2r_1^2}{a^2}-3\gamma+1\right\}\frac{1}{\xi}\right]$$
$$=-\left[\alpha\frac{d\mathrm{G}}{d\alpha}\right]_{\alpha=-\alpha_s}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

where

$$\frac{\left(1-\gamma\frac{\alpha_s^2r_1^2}{a^2}\right)\cos\frac{\alpha_sr_1}{a}+\frac{\alpha_sr_1}{a}\sin\frac{\alpha_sr_1}{a}}{\left(1-\gamma\frac{\alpha_s^2r_0^2}{a^2}\right)\cos\frac{\alpha_sr_0}{a}+\frac{\alpha_sr_0}{a}\sin\frac{\alpha_sr_0}{a}}=\frac{\left(1-\gamma\frac{\alpha_s^2r_1^2}{a^2}\right)\sin\frac{\alpha_sr_1}{a}-\frac{\alpha_sr_1}{a}\cos\frac{\alpha_sr_1}{a}}{\left(1-\gamma\frac{\alpha_s^2r_0^2}{a^2}\right)\sin\frac{\alpha_sr_0}{a}-\frac{\alpha_sr_0}{a}\cos\frac{\alpha_sr_0}{a}}=\xi.$$
$$. \quad . \quad . \quad . \quad . \quad . \quad (54)$$

Also

$$\mathrm{F}(i\alpha)=2i\left[\left\{\left(1-\gamma\frac{\alpha^2r_1^2}{a^2}\right)\cos\frac{\alpha r_1}{a}+\frac{\alpha r_1}{a}\sin\frac{\alpha r_1}{a}\right\}\left\{\sin\frac{\alpha r}{a}-\frac{\alpha r}{a}\cos\frac{\alpha r}{a}\right\}\right.$$
$$\left.-\left\{\left(1-\gamma\frac{\alpha^2r_1^2}{a^2}\right)\sin\frac{\alpha r_1}{a}-\frac{\alpha r_1}{a}\cos\frac{\alpha r_1}{a}\right\}\left\{\cos\frac{\alpha r}{a}+\frac{\alpha r}{a}\sin\frac{\alpha r}{a}\right\}\right]=-\mathrm{F}(-i\alpha),$$
$$. \quad . \quad . \quad . \quad . \quad . \quad (55)$$

and

$$\lim_{\zeta\to 0}.\frac{\mathrm{F}(\zeta)}{\mathrm{G}(\zeta)}=\left[\frac{r^3+(3\gamma-1)r_1^3}{(3\gamma-1)(r_1^3-r_0^3)}\right]. \quad . \quad . \quad . \quad . \quad . \quad (56)$$

Hence, finally,

$$U=\frac{Ar_0^3}{4\mu r^2}\left[\frac{r^3+(3\gamma-1)r_1^3}{(3\gamma-1)(r_1^3-r_0^3)}\right]-\frac{Ar_0^3}{4\mu r^2}\frac{F(i\omega)}{G(i\omega)}\cos\omega t+\frac{A\omega^2r_0^3a^3}{4\mu r^2}\sum_{s=1}^{\infty}\frac{F(i\alpha_s)}{i\alpha_s^3(\omega^2-\alpha_s^2)}$$

$$\left[\frac{\left\{\left(1-\gamma\frac{\alpha_s^2r_0^2}{a^2}\right)\cos\frac{\alpha_sr_0}{a}+\frac{\alpha_sr_0}{a}\sin\frac{\alpha_sr_0}{a}\right\}}{r_0^3\left\{\frac{\gamma^2\alpha_s^2r_0^2}{a^2}-3\gamma+1\right\}\left\{\left(1-\gamma\frac{\alpha_s^2r_1^2}{a^2}\right)\cos\frac{\alpha_sr_1}{a}+\frac{\alpha_sr_1}{a}\sin\frac{\alpha_sr_1}{a}\right\}^2}\right.$$

$$\left.\frac{\left\{\left(1-\gamma\frac{\alpha_s^2r_1^2}{a^2}\right)\cos\frac{\alpha_sr_1}{a}+\frac{\alpha_sr_1}{a}\sin\frac{\alpha_sr_1}{a}\right\}\cos\alpha_st}{-r_1^3\left\{\frac{\gamma^2\alpha_s^2r_1^2}{a^2}-3\gamma+1\right\}\left\{\left(1-\gamma\frac{\alpha_s^2r_0^2}{a^2}\right)\cos\frac{\alpha_sr_0}{a}+\frac{\alpha_sr_0}{a}\sin\frac{\alpha_sr_0}{a}\right\}^2}\right], \quad (57)$$

where γ is given by (47), F by (55), G by (52) and the summation is over the positive zeros of (52).

The radial and hoop stresses are now given by

$$\left.\begin{aligned}\overset{\wedge}{rr}&=(\lambda+2\mu)\frac{\partial U}{\partial r}+2\lambda\frac{U}{r}\\ \overset{\wedge}{\theta\theta}&=\lambda\frac{\partial U}{\partial r}+2(\lambda+\mu)\frac{U}{r}\end{aligned}\right\} . \quad . \quad . \quad . \quad . \quad . \quad (58)$$

LXVI. *Ultra-Violet Dispersion of Air.*

By H. Lowery, D.Sc., F.Inst.P.,
Principal, South-West Essex Technical College, Walthamstow, E. 17 *.

[Received April 7, 1942.]

[Plate I.]

ABSTRACT.

The dispersion of dry air over the range λ6563—λ1822 has been determined by means of a prism used in conjunction with a diffraction grating and the results expressed by the Cauchy formula

$$(n-1)\times10^7=2882{\cdot}2+(13{\cdot}53/\lambda^2\times10^{-8})+(0{\cdot}3064/\lambda^4\times10^{-16}).$$

INTRODUCTION—PREVIOUS WORK.

MANY measurements of the refraction and dispersion of air have been made for the visible and near ultra-violet regions of the spectrum, but relatively little work has been published on the region below λ2600, and only one set of results extends below λ2300.

---

* Communicated by the Author.

The literature has already been fully surveyed by previous writers (1), and attention has been drawn to the somewhat unsatisfactory nature and deficiencies of the data so far tabulated.

Most workers have used some form of interferometer, but as the present investigation was carried out by a prism method no reference will be made here to interferometer methods except in so far as the final results may be compared with those of the prism method.

The prism method has been used for the determination of the refractivity of gases by Biot and Arago, Le Roux, Dulong, von Lang, Carnazzi, McLennan, Statescu, Stuckert, and Hurion. These observers have usually desired to obtain the value of the refractive index for a single wave-length. The deviation of a ray of light passing through a prism of gas is so small that obviously a prism by itself is unsuitable for the study of dispersion. However, when combined with a diffraction grating, it yields a method that is capable of attaining a high degree of accuracy and at the same time gives the dispersion in one sweep over the range of spectrum within the limits of absorption of the elements of the apparatus.

The only previous use of the prism-grating combination in the study of the dispersion of air was that of Kayser and Runge (2) who, in 1893, published results over the range $\lambda 5630$–$\lambda 2360$. They used a 60°-prism of aperture 2×1 cm., the sides being made of quartz, clamped securely so as to withstand great pressure. This was placed between the grating and photographic plate in a Rowland mounting of a 21 ft. concave grating. Spectrograms were taken of an iron arc and tin spark with the prism full of air at atmospheric pressure and then with the air at a pressure of about 10 atmospheres. In this way the spectral lines were displaced by as much as 15 mm. on the photographic plate. Pressure and temperature measurements of the air (not dried) were made and the results reduced to N.T.P. conditions by means of van der Waals's equation.

Not taking into account the various corrections that must be applied to allow for instrumental errors, the theory of the experiment is simple: thus, for a prism at minimum deviation, $n=\sin \frac{1}{2}(A+D)/\sin \frac{1}{2}A$. Since in this case D is small $(n-1)=\frac{1}{2}.D.\cot \frac{1}{2}A$. If d is the shift of a line on the photographic plate and r the distance between the point of divergence of the deviated ray and the plate, $d=rD$, hence

$$(n-1)=\tfrac{1}{2}(d/r)\,.\,\cot \tfrac{1}{2}\,A.$$

The experiment was carried out in two parts, (*a*) with the distance r=about 600 cm.; (*b*) with r about 350 cm., since for wave-lengths below $\lambda 2860$ the smallness of the width of the prism led to a serious reduction in the resolving power of the system with the prism very near the grating. A slightly different prism angle was used in the second part; the final results, however, showed good agreement over the whole spectral range of the experiments.

The quartz sides of the prism were plane but not quite parallel, and were mounted so that the deviation of the light passing through the prism was in the opposite direction to that due to the pressure (10 atmospheres) of the air. The prism covered only one half of the grating, hence two spectra were obtained side by side, one having the deviated lines and the other the undeviated lines for comparison purposes. From the measurements of the deviations of lines of known wave-lengths over ranges of about 200 A.U. mean deviations for mean wave-lengths were calculated, and the final results of the refractive index were given for 8 mean wave-lengths.

EXPERIMENTAL.

In Kayser and Runge's use of the prism-grating system, the prism was placed in the diffracted beam, that is, between the grating and the photographic plate, so that the whole of the diffracted light could not be passed through the prism at one time unless the prism were of small angle or of inconveniently wide aperture. Clearly there is an advantage in using as large a prism angle as possible owing to the correspondingly greater deviation in the beam refracted by the gas, hence, it was decided in the present experiments to work with a wide-angle prism and place it between the slit and the grating.

Preliminary experiments were made with the prism mounted in the path of a parallel beam of light from a quartz lens, the deviated beam from the prism then passing through a diverging lens on its way to the grating, the diverging lens being arranged such that the virtual image of the illuminated slit was at the centre of curvature of the grating. With this arrangement it is easy to show that if A is the angle of the prism, δ the displacement on the photographic plate of the spectral line whose angle of diffraction is θ, ϕ being the angle of incidence on the grating and f the focal length of the diverging lens, then n, the refractive index of the gas is given by $(n-1)=\delta \,.\, \cot \frac{1}{2}A \,.\, \cos \theta/2f \cos \phi$. This arrangement worked well for restricted portions of the spectrum, but would have been troublesome, and probably unsatisfactory, from a practical point of view in a lengthy investigation, owing to the necessity for frequent re-focusing of the two lenses for extensive spectral ranges and was therefore replaced by the following method, in which the lenses were no longer necessary.

Whereas in the above method the prism was placed in a parallel beam of light, now it was placed in the diverging beam from the slit to the grating, precautions being taken in the application of the method that no serious error should arise in regarding the diverging beam as approximately parallel with the central ray as representative. Rays which make an angle with the central ray through the prism will be deviated more than the central ray (assuming minimum deviation position of the prism), and this would have the effect of producing a broadening of the virtual image of the slit and thus giving diffuse lines on the plate. This difficulty was avoided by masking a part of the grating aperture until sufficiently sharp lines were obtained on the photographic plate. Moreover, the prism was placed as far from the slit as possible (that is, as near to the grating as convenient) so as to reduce to the minimum the effect of divergence of the beam meeting the prism.

The final arrangement of the apparatus was as follows: the slit, grating, and photographic plate were mounted on the Rowland circle with the slit at the centre of curvature of the grating. The prism was placed symmetrically between the slit and the grating, as near to the latter as possible—in practice at a distance of about 30 cm. The exact position of the slit was, of course, determined experimentally in order to correct for the refraction of the quartz plates of the prism and bring the spectrum into focus on the photographic plate. Spectra were obtained (*a*) with the prism evacuated using the upper two-thirds of the slit, and (*b*) with the prism full of air using the lower two-thirds of the slit, so that the displacement of the lines on the photographic plate due to refraction by air in the prism was readily measurable. Pl. I. illustrates the appearance on the photographs.

If α is the deviation produced by the prism,

$$\alpha = 2(n-1)\tan\tfrac{1}{2}\mathrm{A}.$$

For the grating $\lambda = b(\sin\theta + \sin\phi)$, where b is the grating space and λ the wave-length of the light; hence

$$d\theta/d\phi = \cos\phi/\cos\theta \text{ numerically, and } d\theta = (\cos\phi/\cos\theta).\alpha d/\mathrm{D},$$

where d is the distance between the slit and prism and D the diameter of the Rowland circle. Now δ, the displacement of a spectral line on the plate, $=\mathrm{D}.d\theta$, where $d\theta$ is the corresponding change in the angle of diffraction; therefore, $\delta = \alpha.d.\cos\phi/\cos\theta$ and, finally,

$$(n-1) = \tfrac{1}{2}.\delta.\cos\theta.\cot\tfrac{1}{2}\mathrm{A}/d.\cos\phi,$$

which gives the relation between the refractive index, the displacement on the photographic plate, and the constants of the apparatus. As the investigation is concerned with dispersion only (see below), it is not necessary to evaluate the latter constants.

The prism was constructed from a suitably drilled, massive brass casting by attaching two quartz plates as sides, the whole forming a hollow 120°-prism with provision for connexion to exhaust pumps, gas apparatus, and thermometers. This was mounted about 30 cm. from a Rowland concave grating of radius 6 ft. and 14,438 lines per inch. Very thin Ilford photographic plates (15 in. × 2 in.) were used, being held in the arc of the Rowland circle by a specially constructed plate carrier which enabled satisfactory measurements to be obtained from spectral lines throughout the entire length of the plate. Much preliminary work was needed in order to achieve the optimum adjustments of the apparatus for sharp spectral lines from light that had passed through the prism and been reflected from the grating. Ultimately it was found possible to obtain spectral lines as sharp with the prism in position as when photographed without it. Spectrograms were obtained (*a*) with the prism evacuated, and (*b*) with the prism containing dry air, but not deprived of carbon dioxide, at a pressure of the order one to three atmospheres. Accurate measurements of temperature and pressure were recorded.

For most of the experiments an iron arc was used as source with the mercury lamp and other metal arcs as additional sources for special lines. Filters were necessary for the control of exposure in certain parts of the spectral range.

The displacements of the spectral lines at particular wave-lengths were determined on a dividing engine and checked from the deviations of neighbouring lines, using a graphical method for comparison of the deviations over a short range of wave-lengths embracing any selected wave-length. In reducing the final results, experimentally determined corrections were applied to allow for the deformation of the gelatine in the bending of the photographic plates and for the departure of the ends of the plate from perfect fit along the Rowland circle.

It is easy to show that the variation of the refractive index of the quartz of the prism over the spectral range of the experiments cannot introduce an error greater than 1 part in 3000 in the value of $(n-1)$, and, in fact, this error was neglected in computing the results.

Values of $(n-1)$ were reduced to N.T.P. conditions.

RESULTS AND DISCUSSION.

As the object of the experiments was to determine the dispersion of air, no attempt was made to obtain an absolute value of refractive index with the present apparatus; accordingly a previous interferometer measurement made by the author ⁽⁹⁾ was used as a basis for the dispersion curve, viz. $(n-1)\times 10^7=2931$ for λ5461 (Hg), the air being dried but not deprived of carbon-dioxide (*cf.* Meggers and Peters, *loc. cit.*⁽¹⁾, p. 723). The values of $(n-1)$ for other wave-lengths were derived from this and are given in Table I. In addition to the study of the ultra-violet dispersion it was found convenient to include much of the visible spectrum as well, and the results are given for comparison with the interferometer measurements.

The observed values lead to the Cauchy formula

$$(n-1)\times 10^7=2882{\cdot}2+(13{\cdot}53/\lambda^2\times 10^{-8})+(0{\cdot}3064/\lambda^4\times 10^{-16}),$$

from which the values in the third column have been calculated.

The value of the index for sodium light, λ5890, is given by

$$(n-1)\times 10^7=2923{\cdot}3.$$

It will be seen from the second and fourth columns that the results for the visible region obtained by the prism and interferometer methods respectively agree satisfactorily.

The values of $(n-1)$ for the wave-lengths 1942 (Hg), and 1850 (Hg) were not used in calculating the formula, as it was only possible to measure the corresponding displacements to an accuracy of 1 part in 500 owing to the broadness of the lines on the photographic plates.

The lead line, λ1787, was just discernible on the plates, but was so diffuse as to render measurements unreliable. It may be noted that, as a rule, measurements were in the main made on faint sharp lines rather than on the broad lines of great intensity. In most cases the lines were photometered in order to secure that the displacements should be measured accurately from centre to centre.

TABLE I.

Dispersion of Air (Dry).

Wave-length in A.U.	Values of $(n-1)\times 10^7$ reduced to N.T.P. conditions.		
	Observed.	Calculated from Cauchy formula.	Previously observed by interferometer.
6563 Hα	2914	2914	2914
6438 Cd	2916	2917	2916
5461 Hg	2931	2931	2931
5209 Ag	2937	2936	2937
5086 Cd	2939	2939	2940
4861 Hβ	2945	2945	2946
4800 Cd	2948	2947	2948
4358 Hg	2962	2962	
4047 Hg	2976	2976	
3906 Hg	2983	2984	
3650 Hg	3002	3001	
3387 Fe	3025	3024	
3200 Fe	3044	3043	
3126 Hg	3053	3053	
3018 Fe	3067	3067	
2929 Fe	3081	3081	
2413 Fe	3204	3205	
2338 Fe	3230	3232	
2178 Fe	3303	3304	
1942 Hg	3460*	3456	
1850 Hg	3543*	3549	
1822 Pb	3569	3568	

* Experimental values not used in determining the formula (see text).

Table II. is an extension of dispersion data collected in Landolt and Bornstein's Tables and shows the difference between the refractivities at certain selected wave-lengths and the refractivity for the sodium line λ5890, for comparison purposes. Down to λ2600 the dispersion curves of the various investigators are reasonably similar, though in some cases the accuracy of the dispersion measurements is far from being as good as 0·1 per cent. Beyond λ2600 little work has been done, and there are only two sets of results beyond λ2300, viz., those of Traub and the present.

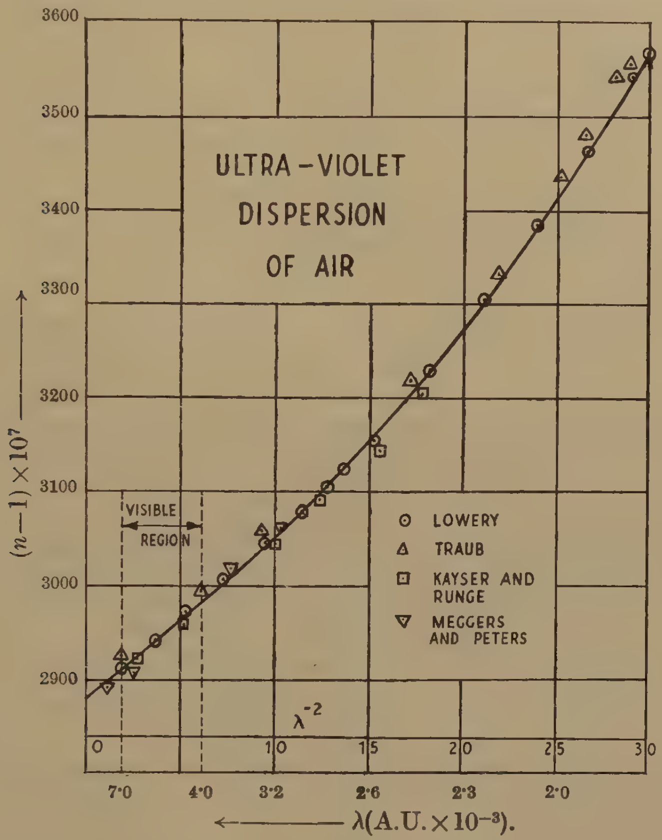
TABLE II.

$(n - n_{\lambda 5890}) \times 10^7$.

Wave-length in A.U.	Present work.	Traub.(3)	Kayser and Runge.(2)	Meggers and Peters.(1)	Howell.(4)	Dickey.(5)	Koch.(6)	Quarder.(7)	Rent-schler.(8)
5890	0	0	0	0	0	0	0	0	0
5460	7	9	7	7	—	5	7	—	6
5360	9	—	9	9	—	7	—	—	—
5050	16	—	15	16	12	15	—	16	—
4860	22	—	21	21	—	23	—	29	—
4700	26	—	25	26	—	26	—	—	—
4360	39	42	39	39	28	41	41	42	32
4050	53	57	52	54	—	55	56	—	48
3650	78	—	76	80	—	81	81	—	73
3340	105	110	103	108	111	108	—	109	112
2960	153	159	151	159	155	156	157	150	—
2620	220	225	219	236	231	242	—	210	—
2360	289	302	294	321	—	—	—	—	—
2030	471	474	—	—	—	—	—	—	—
1850	632	642	—	—	—	—	—	—	—

The dispersion curves of Kayser and Runge and the present work lie very close together along their whole length (*i. e.*, down to λ2360, the limit of the former work), and the differences are insignificant, even taking into account the reductions to be applied to Kayser and Runge's results in order to allow for the fact that their experiments were made with moist air. The curve of Meggers and Peters is practically coincident with that of the present work as far as λ4300 and then begins to steepen slightly with respect to the latter.

Traub's results extend over the same range as the present, and the differences are too small to be of any consequence ; in fact, taking into account the difficulties of obtaining reliable results over a wide range, all four curves just referred to are remarkably concordant (see figure, which shows values of $(n-1)\times 10^7$ plotted against λ^{-2}).



The work of Meggers and Peters (range λ8688—λ2392 for N.T.P. conditions) has been described as classical (1c), especially in view of the fact that it provided detailed tables for the reduction of wave numbers from " air " to " vacuum " conditions for spectroscopic purposes. For the range of their work, viz., λ8688—λ2392 (0° C. and 760 mm.), Meggers and Peters made 277 index measurements for 122 different wave-lengths with a probable error of 3·8 in a single value of $(n-1)\times 10^7$. In the present work over 1800 measurements of displacements of spectral lines were made over the range λ6563—λ1822, and the Cauchy formula was calculated from the reduced observations at 20 selected points in the range. The probable error is 0·5. It may be noted that the International Critical Tables reproduce only one list of refractivities of air, viz., the *uncorrected* experimental observations of Meggers and Peters for dry air at 0° C., 15° C., and 30° C. respectively, whereas it is customary in practical applications to employ results calculated from dispersion formulæ representing smooth curves based on the observed data.

The Cauchy formulæ of Meggers and Peters, Kayser and Runge, and of the present work are as follows :—

Meggers and Peters :

$$(n-1)\times 10^7 = 2875{\cdot}66 + (13{\cdot}412/\lambda^2\times 10^{-8}) + (0{\cdot}3777/\lambda^4\times 10^{-16}).$$

Kayser and Runge :

$$(n-1)\times 10^7 = 2881{\cdot}7 + (13{\cdot}16/\lambda^2\times 10^{-8}) + (0{\cdot}316/\lambda^4\times 10^{-16}).$$

Present Work :

$$(n-1)\times 10^7 = 2882{\cdot}2 + (13{\cdot}53/\lambda^2\times 10^{-8}) + (0{\cdot}3064/\lambda^4\times 10^{-16}).$$

Acknowledgments.

I wish to acknowledge the help given me in the earlier experiments by the late Mr. Louis Frith, B.Sc., M.Sc.Tech., a brilliant young physicist who was tragically killed in a motor cycling accident. I also wish to thank Mr. Derek J. Price, my research assistant, for assistance with the numerous calculations involved in the work ; Mr. B. Hyde Harvey, Borough Treasurer of Leyton, for access to a Mercedes Euklid electric calculating machine ; and the Government Grant Committee of the Royal Society for the loan of certain apparatus.

References.

(1) (*a*) Cuthbertson, C. and M., Proc. Roy. Soc. lxxxiii. p. 151 (1910).
 (*b*) Meggers and Peters, 'Scientific Papers of the Bureau of Standards,' no. 327, p. 697 (1918).
 (*c*) Barrell and Sears, Phil. Trans. A., Roy. Soc. ccxxxviii. p. 1 (1939).
(2) Kayser and Runge, *Abhandl. d. Berl. Akad. d. Wissensch.* p. 3 (1893).
(3) Traub, *Ann. d. Phys.* lxi. p. 533 (1920).
(4) Howell, Phys. Rev. vi. p. 81 (1915).
(5) Dickey, Astroph. Jour. xlv. p. 188 (1917).
(6) Koch, *Nova Acta Upsala*, ii. p. 4 (1909).
(7) Quarder, *Ann. d. Phys.* lxxiv. p. 255 (1924).
(8) Rentschler, Astroph. Jour. xxviii. p. 345 (1908).
(9) Lowery, Proc. Phys. Soc. xl. p. 23 (1927).

LXVII. *Dimensions.*

By Prof. W. WILSON, F.R.S.[*]

[Received July 11, 1942.]

It is with some hesitation that I venture to break into the dimensional controversy a second time. Dr. Brown[†] appears to have slipped out of my net; but I believe he is still within it somewhere. His apparent escape is achieved by writing the dimensions of λ and C in the formula

$$\tau=2\pi\sqrt{\lambda C} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

as follows:

$$\left.\begin{aligned}[\lambda]&=[L],\\ [C]&=[T^2L^{-1}],\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

since the product, λC, has the dimensions $[T^2]$. I would agree with

$$\left.\begin{aligned}[\lambda]&=[\mu L],\\ [C]&=[T^2L^{-1}\mu^{-1}],\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

I shall try to meet him by adopting the electromagnetic system of units, since he insists that equations (2) are based on this system. In this system, indeed in the electrostatic system too, equation (1) takes the form

$$\tau=\sqrt{\mu K}.\theta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where θ is, for a given circuit, a quantity which can only vary with the unit of length chosen. It has, in fact, the dimensions of a length. What I mean by this is made clear, I think, when I say that if I were to double the unit of length the numerical value of θ would be halved[‡]; but it would be unaffected by changing any other unit. The meanings of μ and K are well defined in both the e.s. system and in the e.m. system. For air we have, approximately,

$$K=1,$$
$$\mu=1/9\times10^{-20},$$

in the e.s. system, while

$$K=1/9\times10^{-20},$$
$$\mu=1,$$

approximately, in the e.m. system. Now it completely transcends my limited understanding to discern the error in the following inference from (4):

---

[*] Communicated by the Author.

[†] G. B. Brown, "A Note on Prof. Wilson's Dimensions of Physical Quantities," Phil. Mag. xxxiii. p. 367 (1942).

[‡] And it follows, necessarily (if I retain the same unit of time), that the value of $\sqrt{\mu K}$ would be doubled. Should this happen to a dimensionless quantity?

$$[\tau]=[\mu^{\frac{1}{2}}.K^{\frac{1}{2}}L]. \qquad (5)$$

May I assure Dr. Brown that I am not attempting to mix his theory (which indeed I do not understand) with Maxwell's. The formula (1) stands quite independently of *any* theory—though it was originally obtained in a theoretical way. The λ and C are well-defined quantities, and the formula expresses an experimental fact. The expression for C, in the case of a parallel plate condenser for example, is

$$\frac{K \times \text{area}}{4\pi d},$$

both in the old-fashioned e.s. units and in the old-fashioned e.m. units. I say "old-fashioned" to avoid confusion with Lorentz–Heaviside systems. Physicists do actually speak of units of capacity: the e.m. unit, the e.s. unit, the farad, etc. This necessarily implies units for dielectric constant. These are the dielectric constant of free space in the e.s. system and a unit 9×10^{20} times as great in the e.m. system. Similar things may be said about permeability. Indeed, if I were to double the unit of length, without changing that of time, then in the e.m. system in which $\mu=1$ (for free space) the numerical value of K would change from $1/9 \times 10^{-20}$ to $4/9 \times 10^{-20}$. And yet, according to Brown, it is unitless. I fail to see the relevancy, or point, in "K is not defined from Prof. Wilson's equation (6); if it were we should not only never find the dimensions of electric charge, we should never find the unit." My equation (6) is

$$F=AQ_1Q_2/Kr^2, \qquad (6)$$

in which A, in the old-fashioned systems, is unity. This equation, or its consequence in the form of the above expression for the capacity of a condenser, certainly can be made use of in defining K; but for this purpose it has to be gingered up by the experimental fact that the capacity of my parallel plate condenser is altered when I exchange one dielectric medium for another. Were this not the case I should fail to find the K. I confess frankly my inability to express the dimensions of the Q in (6) in terms of those of Mass, Length and Time *only*; but I find it almost childishly simple to "find the unit" of electric charge. It was not my intention to give the impression that I considered it to be one of the functions of (6) to provide definitions of K or Q or both, or even of their units. I could easily provide these definitions, *even if I were to become totally oblivious of equation* (6); but, of course, this equation did suggest the old-fashioned e.s. unit.

I fear that Dr. Brown and I will never understand one another.

Prof. Dingle's* purpose is "to relate the subject" (of dimensions) "to more fundamental and universally acceptable ideas and thus establish a set of principles to which the questions under discussion may be referred for settlement." He then proceeds to twelve definitions and postulates. He tries to make them "clear and convincing" and is of opinion that "each taken by itself will probably command immediate assent." Yet there are some, and seemingly the most important of them, that I do not understand.

---

* H. Dingle, "On the Dimensions of Physical Magnitudes," Phil. Mag. xxxiii. p. 321 (1942).

Whatever may be the value of his fundamental ideas, definitions and postulates, my confidence in them is severely shaken by certain inaccuracies and misconceptions which occur in his paper. I read, for instance, "In the formula

$$t=2\pi\sqrt{\frac{l}{g}}\text{''}$$

which he numbers (13). "for the time of swing of a simple pendulum, for instance, we may, if we wish, speak of 2π radians, but if we decide to express angles in degrees, the coefficient does not become $2\pi\times 57{\cdot}3$." This statement surprises me! Ever since I first felt that I had understood the credentials of this formula, I appear to have laboured under the misapprehension that the 2π in it has nothing whatever to do with angles or radians. Surely the solution of the relevant differential equation is expressed in terms of the function

$$e^{i\sqrt{\frac{g}{l}}\,t}$$

(or of some equivalent function) and $e^{i\phi}$ is periodic in ϕ, its period being the *number* 2π; hence the familiar formula for the period.

Further on we come to the measurement of temperature and read "Progress in physical theory, supported by all the experimental evidence available, indicated that the temperature of a body as so defined" (perfect gas scale) "was proportional to the mean kinetic energy (kinetic energy per unit mass) of the molecules of the body." So the author arrives at the equation, which he numbers (14),

$$\theta=\alpha\frac{\mathrm{E}}{\mathrm{M}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (7)$$

where α, as he maintains, is a constant. The proportionality statement may be accepted; but he proceeds "But then equation (14) became tacitly adopted as a *definition* of temperature, α, of course, having a specific numerical value assigned to it, and, therefore, becoming dimensionless." Here I meet with difficulties. Prof. Dingle's language suggests to me that those physicists, or astrophysicists, who calculate temperatures from this "tacitly adopted" definition, are guilty of some lack of precision. Does he mean by "specific numerical value" that the *same* value of α is employed by them in all cases? I feel that they are too clever for this. They would probably only use such a definition in circumstances where the simplest form of kinetic gas theory could be safely adopted, and they would probably be aware that the gas constant per unit mass, as well as the number of degrees of freedom of the molecule, vary from one gas to another. They would probably feel on safer ground if they defined the temperature as proportional to the average kinetic energy *of translation* per *gramme molecule*. Then they might indeed adopt a "specific numerical value" for the constant of proportionality, but they would express it, or its reciprocal, in terms of ergs per gramme molecule per degree, or in similar terms. They would hardly think of it as "dimensionless." How could they? I may be misinterpreting Prof. Dingle's meaning; but he is just the one natural philosopher who cannot afford to use language which is lacking in precision or meaning. Has he not said, in connexion with this very subject of the definition of temperature: "Discussions on dimensions thus become richly endowed with meaningless verbosity . . . . ?"

Before leaving Prof. Dingle's section on the measurement of temperature, I must make some brief observations on the following passage: "In interstellar matter, for example, the kinetic energy per unit mass corresponds to a temperature of some thousands of degrees, but there is every reason to believe that a thermometer placed there would read about 3° absolute." I understand this statement, I believe, and I will refrain from hypercriticism of the concluding observation that "a thermometer placed there would read about 3° absolute." I imagine—and I hope I am not misinterpreting the author—that this passage is intended to bring out actual or supposed discrepancies between different definitions of temperature. We may be reassured. All definitions of temperature I have ever heard of associate it with a kind of equilibrium, often called thermal equilibrium, and identified, in the simple kinetic theory, on which (7) rests, with statistical equilibrium. So the reading "3° absolute," has no meaning, so far as temperature is concerned. The other one "of some thousands of degrees" possibly has a meaning. It may refer to a region sufficiently large and sufficiently remote from stars, for some reasonable approach to statistical equilibrium to exist. The discrepancy would not appear, therefore, to be due to any conflict between different definitions of temperature. Or, better said, the apparent discrepancy is not a real one.

I pass on to Prof. Dingle's section on *Electrical and Magnetic Magnitudes*. Here I find the following definition of electrical charge:

$$Q=NF. \qquad (8)$$

This is his equation (42), Q and F meaning charge and force respectively. From this he infers, rather temerariously as I think,

$$[Q]=[F]. \qquad (9)$$

But I will give him this point; *i. e.*, I am going to assume that (9) is a correct dimensional statement. From it we obtain, as he does,

$$[K]=[FL^{-2}]. \qquad (10)$$

These statements (9) and (10) are numbered (43) and (45) in his paper. Then follows equation (46), namely

$$\left.\begin{aligned}[Q]&=[MLT^{-2}],\\ [K]&=[ML^{-1}T^{-2}].\end{aligned}\right\} \qquad (11)$$

It is hardly necessary to say that K means dielectric constant.

For the sake of the argument I will make one further gift to Prof. Dingle, namely equation (47), or, to be more precise, I will grant him the dimensionless character of H (magnetic field intensity) in equation (47). Hence we get

$$\left.\begin{aligned}[m]&=[\mathrm{MLT}^{-2}],\\ [\mu]&=[\mathrm{ML}^{-1}\mathrm{T}^{-2}],\end{aligned}\right\} \quad . \ . \ . \ . \ . \ . \ . \ . \quad (13)$$

in which m and μ mean pole strength and permeability respectively. He does not, I think, give (13) in his paper; but they are implied by what precedes them.

Then he proceeds to make use of the following equation:

$$\mathrm{H}=\frac{\mu\mathrm{Q}}{rt}, \quad . \ . \ . \ . \ . \ . \ . \ . \quad (14)$$

in which μ is expressly declared to be equivalent to permeability! He is thus enabled to shock both Dr. Brown and me with

$$[\mu\mathrm{K}]=[\mathrm{L}^{-1}\mathrm{T}].$$

The correct expression for H is, of course,

$$\mathrm{H}=\frac{\mathrm{Q}}{rt}\times(\text{some number}), \quad . \ . \ . \ . \ . \ . \quad (15)$$

and—I admit it frankly—it is some relief to me to find that, on replacing the erroneous statement (14) by the correct one, (15), the less startling result,

$$[\mu^{-\frac{1}{2}}\mathrm{K}^{-\frac{1}{2}}]=[\mathrm{LT}^{-1}],$$

emerges. It is easy to see this.

Prof. Dingle and I agree that the H in question has zero dimensions. Hence by (15)

$$[\mathrm{LT}]=[\mathrm{Q}].$$

Therefore $[\mathrm{LT}]=[\mathrm{F}]$, by (9),

and so $[\mathrm{M}]=[\mathrm{T}^3]$.

On substituting in (11) we get

$$[\mathrm{K}]=[\mathrm{TL}^{-1}],$$

and similarly, by (13), we find

$$[\mu]=[\mathrm{TL}^{-1}].$$

In conclusion, I wish to make some more general remarks. I believe I am beginning to understand what Prof. Dingle is groping after when he is referring to definitions of the type

$$\mathrm{Y}=\mathrm{NX}, \quad . \ . \ . \ . \ . \ . \ . \ . \quad (16)$$

in which X is already known, N is *declared* to be a specific number and thus dimensionless, and Y is the "magnitude" that is being defined. This kind of thing is sound—at all events in many cases, some of which were mentioned in my paper—and I am hoping for some measure of agreement between him and me. There will be nothing relativistic about its achievement—he would have to leave his point of view and come over to mine, naturally !

If his equation (14), which I have numbered (7) above, were modified to

$$\theta = \alpha \frac{\text{Average kinetic energy of translation}}{\text{Number of molecules}}, \quad . \quad . \quad . \quad (17)$$

I should agree to accept it as a definition of θ, if a *specific numerical value* were assigned to α, and I should agree with

$$[\theta] = [\text{Energy}].$$

This illustration is given in my paper, and I have preferred to put it in the form of defining Willard Gibbs' modulus to be the temperature or rather a substitute for it. This kind of definition is rendered safe by the existence of the universal constant k, the so-called entropy constant. Prof. Dingle has been unfortunate in his expression of what he seems to mean. He should not have used kinetic energy *per unit mass* as the basis of a definition of temperature. Eddington's

$$\text{Mass of Sun} = 1{\cdot}47 \text{ kilometres}$$

(or whatever it is) represents a legitimate use of definitions of the form (16). Here again we have a universal constant, or rather two of them, involved.

But Prof. Dingle's equation

$$Q = NF,$$

and his definition of H as a dimensionless quantity only confuse, it seems to me, the issues in this discussion. What a pity he did not think of cases like the following :

$$\text{Heat} = N \times \text{Work},$$

in which the *specific numerical value*, $N=1$, identifies heat and energy, or

$$\text{Energy} = N \times \text{Mass},$$

in which, if $N=1$, mass and energy are identified.

Prof. Dingle and Dr. Brown have, both of them, a strong philosophical bent ; may I recommend to their attention the pragmatic philosophy ? I am sure William James would have asked, if he could have read this discussion : What practical difference will it make to physics if we adopt Dingle's attitude to dimensions, or Brown's, or even Wilson's ? If we adopt Dingle's we are likely to get lost in wandering mazes of conflicting definitions and worse ; if we adopt Brown's we shall be in the difficult position of having different, and conflicting, dimensional statements in different systems of units, *e. g.*

$$[C] = [L], \text{ e.s.} \quad \text{and} \quad [C] = [T^2L^{-1}], \text{ e.m.} ;$$

if we adopt Wilson's, we deal with any generally accepted relationship between physical quantities, whether it be a definition or experimental fact or a combination of both, *e. g.*

$$m\frac{d^2r}{dt^2}=\mathrm{Q}_1\mathrm{Q}_2/\mathrm{K}r^2,$$

merely by introducing square brackets, or even by leaving them to be understood, thus

$$[\mathrm{MLT}^{-2}]=[\mathrm{Q}^2\mathrm{K}^{-1}\mathrm{L}^{-2}].$$

This kind of thing is often useful. It is simple and absolutely secure from all possibility of contradiction.

LXVIII. *Symmetric Relations between the Coefficients of Reversed Power Series.*

By W. E. Bleick, Ph.D., U.S. Naval Academy *.

[Received January 9, 1942.]

The relation between the coefficients $b_1, b_2, \ldots, b_n$ of the series $y=x+b_1x^2+b_2x^3+\ldots$ and the coefficient c_n of the reversed series $x=y+c_1y^2+c_2y^3+\ldots$ has been worked out by C. E. Van Orstrand † as far as $n=12$. These relations, which are rather cumbersome for practical use in reverting series, are transformed in the present paper to a comparatively simple recurrent form.

An obvious way to recast Van Orstrand's relations in a recurrent form is to manipulate them in such a way that the coefficients b_i and c_i appear in a symmetric fashion. This symmetrization has been accomplished in terms of the auxiliary quantity $a_i=b_1c_{i-1}+b_2c_{i-2}+\ldots+b_{i-1}c_1$, which is symmetric in the coefficients of the two power series. The method employed was to express b_n+c_n as a polynomial in $a_2, a_3, \ldots, a_n$ with unknown coefficients. The unknown coefficients were then evaluated by expanding the polynomial in terms of $b_1, b_2, \ldots, b_{n-1}$, and equating the coefficients of the various powers of $b_1, b_2, \ldots, b_{n-1}$ to the numerical values given in Van Orstrand's formulæ. The symmetric relations obtained by this process are:—

$$b_1+c_1=0,$$

$$b_2+c_2+2a_2=0,$$

$$b_3+c_3+\frac{5}{2}a_3=0,$$

$$b_4+c_4+3a_4+a_2{}^2=0,$$

---

* Communicated by the Author.

† "Reversion of Power Series," Phil. Mag. (6) xix. p. 366 (1910). The author has changed Van Orstrand's notation slightly by reversing the algebraic signs of b_1, b_2, etc.

$$b_5+c_5+\frac{7}{2}a_5+\frac{7}{2}a_2a_3=0,$$

$$b_6+c_6+4a_6+6a_2a_4+\frac{5}{2}a_3{}^2+4a_2{}^3=0,$$

$$b_7+c_7+\frac{9}{2}a_7+\frac{39}{4}a_2a_5+\frac{15}{2}a_3a_4+\frac{93}{4}a_2{}^2a_3=0,$$

$$b_8+c_8+5a_8+15a_2a_6+\frac{45}{4}a_3a_5+5a_4{}^2+46a_2{}^2a_4+\frac{77}{2}a_2a_3{}^2+29a_2{}^4=0,$$

$$b_9+c_9+\frac{11}{2}a_9+22a_2a_7+\frac{33}{2}a_3a_6+\frac{55}{4}a_4a_5+88a_2{}^2a_5+\frac{77}{4}a_3{}^3+\frac{539}{4}a_2a_3a_4$$
$$+242a_2{}^3a_3=0,$$

$$b_{10}+c_{10}+6a_{10}+31a_2a_8+\frac{47}{2}a_3a_7+19a_4a_6+\frac{35}{4}a_5{}^2+160a_2{}^2a_6+\frac{371}{4}a_3{}^2a_4$$
$$+105a_4{}^2a_2+\frac{947}{4}a_2a_3a_5+536a_2{}^3a_4+\frac{1317}{2}a_2{}^2a_3{}^2+317a_2{}^5=0,$$

$$b_{11}+c_{11}+\frac{13}{2}a_{11}+\frac{169}{4}a_2a_9+\frac{65}{2}a_3a_8+26a_4a_7+\frac{91}{4}a_5a_6+\frac{1105}{4}a_2{}^2a_7$$
$$+\frac{611}{4}a_3{}^2a_5+\frac{273}{2}a_4{}^2a_3+\frac{1625}{4}a_2a_3a_6+\frac{2691}{8}a_2a_4a_5+\frac{9373}{8}a_2{}^3a_5$$
$$+\frac{20709}{8}a_2{}^2a_3a_4+\frac{28067}{8}a_2{}^4a_3+\frac{11583}{16}a_2a_3{}^3=0,$$

and

$$b_{12}+c_{12}+7a_{12}+56a_2a_{10}+\frac{175}{4}a_3a_9+35a_4a_8+\frac{119}{4}a_5a_7+14a_6{}^2+455a_2{}^2a_8$$
$$+\frac{1001}{4}a_3{}^2a_6+63a_4{}^3+\frac{2695}{4}a_2a_3a_7+539a_2a_4a_6+\frac{1981}{8}a_2a_5{}^2$$
$$+\frac{3339}{8}a_3a_4a_5+2449a_2{}^3a_6+\frac{20729}{4}a_2{}^2a_3a_5+2256a_2{}^2a_4{}^2$$
$$+\frac{31273}{8}a_2a_3{}^2a_4+\frac{4455}{16}a_3{}^4+8567a_2{}^4a_4+\frac{108999}{8}a_2{}^3a_3{}^2+4819a_2{}^6=0.$$

A finite difference analysis of the numerical coefficients in these twelve symmetric relations gives the general form of the terms which are linear, quadratic, and cubic in a_m. The coefficient of a_r is $(r+2)/2$. The coefficient of a_ra_s is

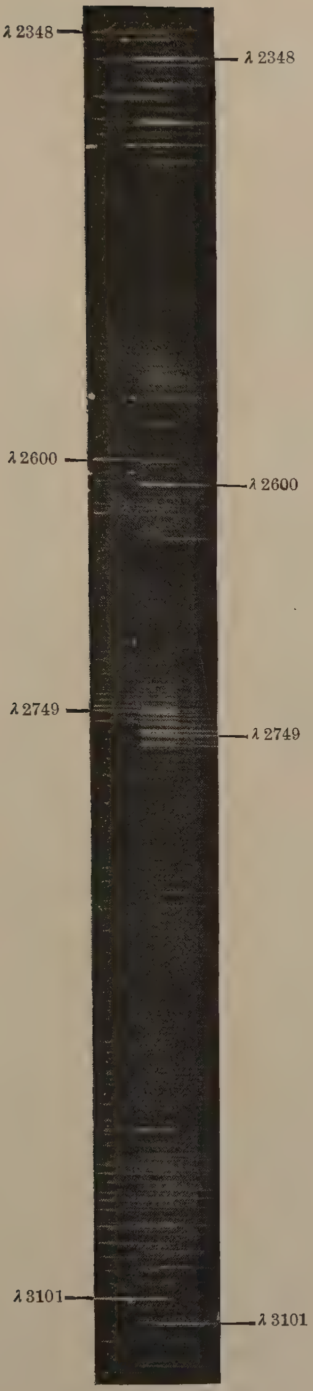
$$\frac{r(r+1)(r+2)+s(s+1)(s+2)}{4(3\,!\,i\,!\,j\,!)},$$

where $i=j=1$ if $r\neq s$, and $i=2$, $j=0$ if $r=s$. The coefficient of $a_ra_sa_t$ is the symmetric sum

$$\Sigma\frac{4r(r+1)(r+2)(r^2+2r+2)+5r(r^3+5r^2+10r+4)(s+t)+5r^2(r+3)(s^2+t^2)}{4(5\,!\,i\,!\,j\,!\,k\,!)},$$

where $i=j=k=1$ if $r\neq s\neq t$, and $i=3$, $j=k=0$ if $r=s=t$, and $i=2$, $j=1$, $k=0$ if $r=s\neq t$.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]


λ 2348
λ 2348
λ 2600
λ 2600
λ 2749
λ 2749
λ 3101
λ 3101

Ultra-violet dispersion of air (typical spectrogram).

LXIX. *On the Equilibrium of Plates.*

By A. C. STEVENSON *.

[Received April 16, 1942.]

Introduction.

HAROLD JEFFREYS has remarked ⁽¹⁾ that the treatment of thin plates in the standard books on elasticity is very complicated. This is even more true of the treatment of thick plates. The present paper gives an alternative compact treatment of the equations of equilibrium of thick or thin plates by making use of the equations of elasticity in complex variables, after paying some initial attention to the specification of stress in a plate. It goes on to consider the semi-inverse method in the theory of plates by the method of complex potentials, and to illustrate the method by examples, some of the results being new.

1. *The Complex Stress Combinations.*

Taking cartesian co-ordinates x, y, Z, we reserve z for the complex variable $x+iy$, except in the designation of stress, since there can be no occasion to confuse z in $\widehat{xz}$, $\widehat{yz}$, $\widehat{zz}$ with the complex variable. Consider, for simplicity, the case of no body force; the three familiar body stress equations

$$\frac{\partial\widehat{px}}{\partial x}+\frac{\partial\widehat{py}}{\partial y}+\frac{\partial\widehat{pz}}{\partial \mathrm{Z}}=0, \quad (p=x,\, y,\, \mathrm{Z}) \quad . \quad . \quad . \quad . \quad (1.1)$$

with the abbreviations

$$\Theta=\widehat{xx}+\widehat{yy}, \quad \Phi=\widehat{xx}-\widehat{yy}+2i\widehat{xy}, \quad \Psi=\widehat{xz}+i\widehat{yz} \quad . \quad . \quad (1.2)$$

become, in terms of variables z, $\bar{z}$, Z, since

$$\frac{\partial}{\partial x}=\frac{\partial}{\partial z}+\frac{\partial}{\partial \bar{z}}, \quad \frac{\partial}{\partial y}=i\left(\frac{\partial}{\partial z}-\frac{\partial}{\partial \bar{z}}\right), \quad \bar{z}=x-iy, \quad . \quad . \quad . \quad (1.3)$$

the two complex equations

$$\frac{\partial\Phi}{\partial z}+\frac{\partial\Theta}{\partial \bar{z}}+\frac{\partial\Psi}{\partial \mathrm{Z}}=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.4)$$

$$\frac{\partial\Psi}{\partial z}+\frac{\partial\bar{\Psi}}{\partial \bar{z}}+\frac{\partial\widehat{zz}}{\partial \mathrm{Z}}=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.5)$$

a bar denoting that i is to be changed to $-i$. Thus, for example, if $\mathrm{F}(z)=i\mathrm{A}z^2$, (A real), $\mathrm{F}(\bar{z})=i\mathrm{A}\bar{z}^2$, $\mathrm{F}(\bar{z})=-i\mathrm{A}\bar{z}^2$. Equation (1.4) is equivalent to two equations, since when we need it, we can write down its conjugate, obtained by changing the sign of i throughout.

---

* Communicated by the Author.

The usefulness of the stress combinations Θ, Φ, Ψ resides in their particular suitability to the problem of the transformation of stress. Thus, if

$$\Theta'=\widehat{nn}+\widehat{ss},\quad \Phi'=\widehat{nn}-\widehat{ss}+2i\widehat{ns},\quad \Psi'=\widehat{nz}+i\widehat{sz}, \quad . \quad . \quad (1.6)$$

where the axes of n and s are merely those of x and y rotated through an angle α, so that

$$z'=n+is=ze^{i\alpha}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.7)$$

we find readily (2) from the consideration of the equilibrium of small wedge elements of the plate, that

$$\Theta'=\Theta,\quad \Phi'=e^{-2i\alpha}\Phi,\quad \Psi'=e^{-i\alpha}\Psi. \quad . \quad . \quad . \quad . \quad . \quad (1.8)$$

In particular, with orthogonal curvilinear co-ordinates ξ, η, given by

$$z=f(\zeta),\quad \zeta=\xi+i\eta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.9)$$

if we write

$$\frac{dz}{d\zeta}=f'(\zeta)=\mathrm{J}e^{i\alpha},$$

where J and α are real, so that

$$\frac{d\zeta}{dz}\frac{d\bar{z}}{d\bar{\zeta}}=e^{-2i\alpha}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.10)$$

it is readily seen that α is the angle made by the tangent at (ξ, η) to the curve $\eta=$const. with the x-axis, and it follows that the transformation to stresses $\widehat{\xi\xi}$, $\widehat{\xi\eta}$, $\widehat{\eta\eta}$, $\widehat{\xi z}$, $\widehat{\eta z}$, $\widehat{zz}$ is given by (1.6), (1.8) and (1.10) on changing n, s to ξ, η respectively.

2. *The Specification of Stress in a Plate.*

If $\mathbf{n}$, $\mathbf{s}$ are unit vectors in the mid-plane $\mathrm{Z}=0$ of a plate (bounded by $\mathrm{Z}=\pm h$), whose directions are respectively normal and tangential to a curve of arc s and $\mathbf{k}$ is the unit vector parallel to the Z-axis, the stress-vector at points of the rectangular element of cross-section of the plate, which is normal to $\mathbf{n}$ and bounded by lines parallel to the Z-axis at the ends of an element of the arc s, is

$$\mathbf{R}_n=\widehat{nn}\mathbf{n}+\widehat{ns}\mathbf{s}+\widehat{nz}\mathbf{k}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.1)$$

$\mathbf{n}$, $\mathbf{s}$, $\mathbf{k}$ being right-handedly related in that order.

The resultant of these stresses over the element, reduced to the mid-point of the arc element ds, gives a certain force and couple resultant. Dividing by ds, and letting ds then tend to zero, we have stress resultants F_{pq} and stress couples G_{pq}, each per unit length of the arc s, given by

$$\mathbf{F}_n=\int_{-h}^{h}\mathbf{R}_n\,d\mathrm{Z}=\mathrm{F}_{nn}\mathbf{n}+\mathrm{F}_{ns}\mathbf{s}+\mathrm{F}_{nz}\mathbf{k}, \quad . \quad . \quad . \quad . \quad . \quad (2.2)$$

$$\mathbf{G}_n=\int_{-h}^{h}\mathrm{Z}[\mathbf{k},\mathbf{R}_n]\,d\mathrm{Z}=\mathrm{G}_{nn}\mathbf{n}+\mathrm{G}_{ns}\mathbf{s}, \quad . \quad . \quad . \quad . \quad . \quad (2.3)$$

so that, more generally, $\mathbf{q}$ being the unit vector in the direction q

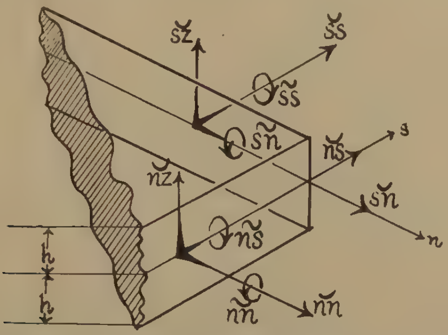
$$\mathrm{F}_{pq}=\int_{-h}^{h}(\mathbf{R}_p,\mathbf{q})\,d\mathrm{Z}=\int_{-h}^{h}\widehat{pq}\,d\mathrm{Z}=2h\overset{\smile}{pq}. \quad . \quad . \quad . \quad (2.4)$$

using Filon's notation $\overset{\smile}{pq}$ for the mean value of $\overset{\frown}{pq}$ across the thickness $2h$ of the plate, *i. e.*, $\overset{\smile}{pq}$ is the stress resultant *per unit area of cross-section normal to p and resolved in the direction q*, and

$$G_{pq}=\int_{-h}^{h} Z(\mathbf{q}[\mathbf{k},\mathbf{R}_p])\,dZ=2h\,\widetilde{pq}, \quad . \quad . \quad . \quad . \quad . \qquad (2.5)$$

introducing a corresponding symbol $\widetilde{pq}$ to denote the stress couple *per unit area of cross-section of the plate normal to* p, *the axis of the couple being in the direction* q.

Stress couples $\widetilde{pq}$, G_{pq} $(p\neq q)$ are termed flexural couples; stress couples $\widetilde{pp}$, G_{pp} are termed torsional couples. If their action on the material behind the element of cross-section is considered, the nomenclature is obviously descriptive (see fig.).



Scheme for nomenclature of stress resultants and stress couples in a plate. From the equality of the cross-shears $\overset{\frown}{ns}$ and $\overset{\frown}{sn}$ we have $\overset{\smile}{ns}=\overset{\smile}{sn}$ and $\widetilde{nn}=-\widetilde{ss}$.

Note that the first letter in the two-letter symbols refers to the element across which the force or couple acts, the second letter to its vector direction. From (2.5)

$$G_{nn}=-\int_{-h}^{h} Z\overset{\frown}{ns}\,dZ, \quad G_{ss}=\int_{-h}^{h} Z\overset{\frown}{sn}\,dZ, \quad . \quad . \quad . \quad . \qquad (2.6)$$

so that the equality of the cross-shears $\overset{\frown}{ns}=\overset{\frown}{sn}$ implies

$$F_{ns}=F_{sn} \quad \text{or} \quad \overset{\smile}{ns}=\overset{\smile}{sn} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2.7)$$

for the stress resultants, and

$$G_{nn}+G_{ss}=0 \quad \text{or} \quad \widetilde{nn}+\widetilde{ss}=0 \quad . \quad . \quad . \quad . \quad . \qquad (2.8)$$

for the stress couples.

Now the definitions (2.4) and (2.5) do not conform, either in symbols or sign conventions, to those of the standard textbooks, and in the writer's opinion much simplicity has been lost by dropping a double suffix stress notation in forming stress resultants and stress couples. In what follows we shall use the stress resultants and couples $\overset{\smile}{pq}$, $\widetilde{pq}$ per unit area of cross-section of the plate rather than those per unit length of a mid-plane arc s, F_{pq}, G_{pq}, as would be more the usual practice. But it seems most unnecessary to restrict the notation $\overset{\smile}{pq}$ to its general use in the case of a plate in a state of generalized plane stress under forces in its own plane, and then change over to another (and inferior) notation when dealing with the same plate under edge couples or pressure over the face $Z=h$.

3. *The Complex Stress and Couple Resultants.*

We use Θ_0, Φ_0, Ψ_0 for the complex stress resultant combinations and Λ, Γ for the complex stress couple combinations, where

$$\Theta_0=\overset{\smile}{xx}+\overset{\smile}{yy}=\frac{1}{2h}\int_{-h}^{h}\Theta\, dZ, \quad \ldots \quad (3.1)$$

$$\Phi_0=\overset{\smile}{xx}-\overset{\smile}{yy}+2i\overset{\smile}{xy}=\frac{1}{2h}\int_{-h}^{h}\Phi\, dZ, \quad \ldots \quad (3.2)$$

$$\Psi_0=\overset{\smile}{xz}+i\overset{\smile}{yz}=\frac{1}{2h}\int_{-h}^{h}\Psi\, dZ, \quad \ldots \quad (3.3)$$

$$\Lambda=\widetilde{xy}-\widetilde{yx}=\frac{1}{2h}\int_{-h}^{h}Z\Theta\, dZ, \quad \ldots \quad (3.4)$$

$$\Gamma=\widetilde{xy}+\widetilde{yx}+2i\widetilde{yy}=\frac{1}{2h}\int_{-h}^{h}Z\Phi\, dZ, \quad \ldots \quad (3.5)$$

so that if the corresponding stress combinations for the co-ordinates n, s of (1.7) are

$$\Theta_0'=\overset{\smile}{nn}+\overset{\smile}{ss}, \quad \Phi_0'=\overset{\smile}{nn}-\overset{\smile}{ss}+2i\overset{\smile}{ns}, \quad \Psi_0'=\overset{\smile}{nz}+i\overset{\smile}{sz}, \quad \ldots \quad (3.6)$$

$$\Lambda'=\widetilde{ns}-\widetilde{sn}, \quad \Gamma'=\widetilde{ns}+\widetilde{sn}+2i\widetilde{ss}, \quad \ldots \quad (3.7)$$

then from (1.8) and (3.1)–(3.5) the equations of transformation of the stress resultants and stress couples are

$$\Theta_0'=\Theta_0, \quad \Phi_0'=e^{-2i\alpha}\Phi_0, \quad \Psi_0'=e^{-i\alpha}\Psi_0, \quad \Lambda'=\Lambda, \quad \Gamma'=e^{-2i\alpha}\Gamma. \quad (3.8)$$

4. *The Complex Identical Relations.*

From (1.4) and (1.5) we have, using (3.1)–(3.5)

$$\frac{\partial\Phi_0}{\partial z}+\frac{\partial\Theta_0}{\partial\bar{z}}+\frac{1}{2h}\Big\{\Psi\Big\}_{-h}^{h}=0, \quad \ldots \quad (4.1)$$

$$\frac{\partial\Psi_0}{\partial z}+\frac{\partial\bar{\Psi}_0}{\partial\bar{z}}+\frac{1}{2h}\Big\{\overset{\smile}{zz}\Big\}_{-h}^{h}=0, \quad \ldots \quad (4.2)$$

$$\frac{\partial\Gamma}{\partial z}+\frac{\partial\Lambda}{\partial\bar{z}}+\frac{1}{2h}\Big\{Z\Psi\Big\}_{-h}^{h}-\Psi_0=0, \quad \ldots \quad (4.3)$$

equivalent to the relations usually termed the identical relations between the stress resultants and stress couples, which we might therefore term the complex identical relations.

5. *Curvature of the Mid-Plane.*

Suppose the indicatrix of the deflected mid-plane $Z=0$ is

$$\kappa_{xx}x^2+2\kappa_{xy}xy+\kappa_{yy}y^2=\text{const.},$$

where $\kappa_{pq}=\partial^2 w_0/\partial p\,\partial q$, w_0 being the Z-deflexion of the mid-plane $Z=0$. In complex co-ordinates this is

$$\beta\bar{z}^2+2\gamma z\bar{z}+\bar{\beta}z^2=\text{const.},$$

where

$$\beta=4\frac{\partial^2 w_0}{\partial\bar{z}^2}=\kappa_{xx}-\kappa_{yy}+2i\kappa_{xy}, \qquad (5.1)$$

$$\gamma=4\frac{\partial^2 w_0}{\partial z\,\partial\bar{z}}=\kappa_{xx}+\kappa_{yy}. \qquad (5.2)$$

The change of axes (1.7) gives the equation of the indicatrix in the new co-ordinates z', $\bar{z}'$ with

$$\beta'=\beta e^{-2i\alpha}, \quad \gamma'=\gamma. \qquad (5.3)$$

Note particularly that these equations of transformation are of the same form as for Φ_0, Θ_0, Γ and Λ in (3.8). If the new axes are principal axes of curvature, $\kappa_{ns}=0$ and with the usual notation $\kappa_{nn}=\kappa_1$ and $\kappa_{ss}=\kappa_2$, then we have

$$\beta'=\kappa_1-\kappa_2, \quad \gamma'=\kappa_1+\kappa_2. \qquad (5.4)$$

6. *Elementary Plate Theory.*

The starting point in the elementary plate theory, via the usual argument from Hooke's law, is that, if a plate is bent into a cylindrical surface of small curvature κ_1, the necessary couples are

$$\widetilde{ns}=-\mathrm{P}\kappa_1, \quad \widetilde{sn}=\eta\mathrm{P}\kappa_1, \quad \widetilde{ss}=0, \qquad (6.1)$$

where η is Poisson's ratio and P is an elastic constant or plate modulus ($2h$P is Love's plate modulus D, see 'Mathematical Theory of Elasticity,' by A. E. H. Love, 4th ed., 1927. p. 464). The superposition of this pure flexure on another about the perpendicular axis, of amount κ_2, namely

$$\widetilde{ns}=-\eta\mathrm{P}\kappa_2, \quad \widetilde{sn}=\mathrm{P}\kappa_2, \quad \widetilde{ss}=0 \qquad (6.2)$$

gives

$$\Gamma'=-(1-\eta)\mathrm{P}\beta', \quad \Lambda'=-(1+\eta)\mathrm{P}\gamma'. \qquad (6.3)$$

Hence, from (3.8) and (5.4), transformation to other axes gives

$$\Gamma=-(1-\eta)\mathrm{P}\beta, \quad \Lambda=-(1+\eta)\mathrm{P}\gamma. \qquad (6.4)$$

If we assume that for a thin plate these are consistent with the assumptions

$$\Psi=0 \text{ over } Z=\pm h, \quad \widehat{zz}=-p \text{ over } Z=h, \quad \widehat{zz}=0 \text{ over } Z=-h, \qquad (6.5)$$

then the complex identical relations (4.2) and (4.3) give

$$\frac{\partial\Psi_0}{\partial z}+\frac{\partial\bar{\Psi}_0}{\partial\bar{z}}=\frac{p}{2h}, \quad \frac{\partial\Gamma}{\partial z}+\frac{\partial\Lambda}{\partial\bar{z}}=\Psi_0, \qquad (6.6)$$

whence

$$\frac{\partial^2\Gamma}{\partial z^2}+\frac{\partial^2\bar{\Gamma}}{\partial\bar{z}^2}+2\frac{\partial^2\Lambda}{\partial z\,\partial\bar{z}}=\frac{p}{2h}. \qquad (6.7)$$

Then from (5.1), (5.2) and (6.4) we find at once

$$16\mathrm{P}\frac{\partial^4 w_0}{\partial z^2\,\partial\bar{z}^2}=-\frac{p}{2h} \qquad (6.8)$$

or

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)^2 w_0=-\frac{p}{2h\mathrm{P}}=-\frac{p}{\mathrm{D}} \qquad (6.9)$$

in the usual notation.

The solution of (6.8), remembering that w_0 is real, is, when p is constant,

$$w_0=\bar{z}\Omega(z)+z\bar{\Omega}(\bar{z})+\omega(z)+\bar{\omega}(\bar{z})+\frac{3}{16}\frac{p}{h^3}z^2\bar{z}^2 \quad . \quad . \quad . \quad (6.10)$$

in terms of two complex potentials $\Omega(z)$, $\omega(z)$, but we shall not pursue the search for $\Omega(z)$, $\omega(z)$ for any particular boundaries any further here, as we are more concerned to show that the corresponding exact semi-inverse solutions for thick or moderately thick plates are intrinsically no more difficult to find.

7. *The Semi-Inverse Method in the Theory of Plates.*

The above treatment of thin plates, besides being immediately useful when the boundaries are members of orthogonal curvilinear families of curves, has the advantage of leading on naturally to the treatment of thick plates by the semi-inverse method, where we assume simple forms for $\widehat{zz}$ and Ψ satisfying the plane boundary conditions (6.5) (taking the case $p=$const. for brevity), and then solve for the remaining stress combinations Θ, Φ and the displacements w, $D=u+iv$ from the complex body-stress equations (1.4) and (1.5) and the stress-strain relations. We shall show that this can be done in terms of four complex potentials, of which two are concerned with the mid-plane Z-deflexion. We thus get complete information as to the state of stress and displacement in the plate, at very small extra cost in labour above that in the elementary theory, and moreover we get some light from the exact solution of the semi-inverse method as to the compatibility of the equations (6.4) and (6.5) of the elementary theory.

The stress-strain relations in complex form give the following stress-displacement equations:—

$$\Phi=4\mu\frac{\partial D}{\partial \bar{z}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7.1)$$

$$\Psi=\mu\frac{\partial D}{\partial Z}+2\mu\frac{\partial w}{\partial \bar{z}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (7.2)$$

$$(1-2\sigma)\Theta-2\sigma\widehat{zz}=2\mu\left(\frac{\partial D}{\partial z}+\frac{\partial \bar{D}}{\partial \bar{z}}\right), \quad . \quad . \quad . \quad . \quad . \quad . \quad (7.3)$$

$$(1-\sigma)\widehat{zz}-\sigma\Theta=2\mu\frac{\partial w}{\partial Z}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7.4)$$

Equivalent equations have been given and used by G. Kolossof ⁽³⁾. The elastic constant σ is related to Poisson's ratio η and the Lamé constants λ and μ by the relations

$$(1-\sigma)(1+\eta)=1, \quad \eta=\lambda/2(\lambda+\mu). \quad . \quad . \quad . \quad . \quad . \quad . \quad (7.5)$$

Further, it is well known ⁽⁴⁾ that the combination of the body-stress equations (1.1) and the stress-strain relations is equivalent to the set of equations due to Beltrami,

$$\nabla^2\widehat{pq}+\frac{\partial^2\chi}{\partial p\,\partial q}=0, \quad (p,\, q=x,\, y,\, Z), \quad . \quad . \quad . \quad . \quad (7.6)$$

where $$(1+\eta)\chi=\Theta+\widehat{zz}. \qquad (7.7)$$

From these can be deduced readily the equivalent complex forms

$$\nabla^2\Theta+4\frac{\partial^2\chi}{\partial z\,\partial\bar{z}}=0, \qquad (7.8)$$

$$\nabla^2\Phi+4\frac{\partial^2\chi}{\partial\bar{z}^2}=0, \qquad (7.9)$$

$$\nabla^2\Psi+2\frac{\partial^2\chi}{\partial\bar{z}\,\partial Z}=0, \qquad (7.10)$$

$$\nabla^2\widehat{zz}+\frac{\partial^2\chi}{\partial Z^2}=0, \qquad (7.11)$$

where $$\nabla^2=\nabla_1{}^2+\frac{\partial^2}{\partial Z^2}, \quad \nabla_1{}^2=4\frac{\partial^2}{\partial z\,\partial\bar{z}}. \qquad (7.12)$$

From (7.7), (7.8), (7.11), and (7.12) we have at once

$$\nabla^2\widehat{zz}=\nabla_1{}^2\chi=-\frac{\partial^2\chi}{\partial Z^2}. \qquad (7.13)$$

8. *The Stresses in Terms of Complex Potentials.*

From (6.5) and (1.5), $\partial\widehat{zz}/\partial Z=0$ over $Z=\pm h$, and we construct the simplest polynomials in Z which satisfy this and (6.5), namely

$$\widehat{zz}=\frac{p}{4h^3}(Z^3-3h^2Z-2h^3), \quad \Psi=\psi(z,\bar{z})\{Z^2-h^2\}. \qquad (8.1)$$

From (7.13) and (8.1)

$$\chi=\chi_0(z,\bar{z})+Z\chi_1(z,\bar{z})-\frac{p}{4h^3}Z^3, \qquad (8.2)$$

whence $$\nabla_1{}^2\chi_0=0, \quad \nabla_1{}^2\chi_1=\frac{3p}{2h^3}.$$

From (7.7) and (8.1)

$$\Theta=\theta_0+\theta_1 Z+\theta_3 Z^3, \qquad (8.3)$$

where $$\theta_0=(1+\eta)\chi_0+\tfrac{1}{2}p, \quad \theta_1=(1+\eta)\chi_1+\frac{3}{4}\frac{p}{h} \qquad (8.4)$$

and $$\theta_3=-(2+\eta)\frac{p}{4h^3}. \qquad (8.5)$$

So $$\nabla_1{}^2\theta_0=0, \quad \nabla_1{}^2\theta_1=(1+\eta)\frac{3p}{2h^3}$$

or $$\theta_0=\tfrac{1}{2}\{\Omega_0'(z)+\bar{\Omega}_0'(\bar{z})\} \qquad (8.6)$$

$$\theta_1=\tfrac{1}{2}(1+\eta)\left\{\Omega_1'(z)+\bar{\Omega}_1'(\bar{z})+\frac{3}{4}\frac{p}{h^3}z\bar{z}\right\}, \qquad (8.7)$$

introducing two complex potentials $\Omega_0(z)$, $\Omega_1(z)$.

From (7.10) and (8.1) we then have

$$\Psi=-\frac{1}{2}\left\{\bar{\Omega}_1''(\bar{z})+\frac{3pz}{4h^3}\right\}(Z^2-h^2), \qquad (8.8)$$

and (1.4) then leads to

$$\Phi=f(\bar{z}, \mathrm{Z})-\tfrac{1}{2}z\bar{\Omega}_0''(\bar{z})+\tfrac{1}{2}(1-\eta)\left\{z\bar{\Omega}_1''(\bar{z})+\frac{3pz^2}{8h^3}\right\}\mathrm{Z}.$$

From (7.9) we next find

$$\frac{\partial^2 f}{\partial \mathrm{Z}^2}(\bar{z}, \mathrm{Z})=2\sigma\bar{\Omega}_0'''(\bar{z})-2(2-\eta)\bar{\Omega}_1'''(\bar{z})\mathrm{Z},$$

leading to

$$f(z, \mathrm{Z})=-\tfrac{1}{2}\bar{\omega}_0''(\bar{z})+\tfrac{1}{2}(1-\eta)\bar{\omega}_1''(\bar{z})\mathrm{Z}+\sigma\bar{\Omega}_0'''(\bar{z})\mathrm{Z}^2-\tfrac{1}{3}(2-\eta)\bar{\Omega}_1'''(\bar{z})\mathrm{Z}^3$$

on integrating and introducing two further complex potentials $\omega_0(z)$, $\omega_1(z)$, or

$$\Phi=\phi_0+\phi_1\mathrm{Z}+\phi_2\mathrm{Z}^2+\phi_3\mathrm{Z}^3, \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.9)$$

where

$$\phi_0=-\tfrac{1}{2}\{z\bar{\Omega}_0''(\bar{z})+\bar{\omega}_0''(\bar{z})\}, \quad \phi_2=\sigma\bar{\Omega}'''(\bar{z}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.10)$$

$$\phi_1=\tfrac{1}{2}(1-\eta)\left\{z\bar{\Omega}_1''(\bar{z})+\bar{\omega}_1''(\bar{z})+\frac{3pz^2}{8h^3}\right\}, \quad \phi_3=-\tfrac{1}{3}(2-\eta)\bar{\Omega}_1'''(\bar{z}). \quad (8.11)$$

Hence the complete solution for the stresses depends upon the two pairs of complex potentials $\Omega_0(z)$, $\omega_0(z)$, $\Omega_1(z)$, $\omega_1(z)$. Separation of the stresses from the complex combinations is simple, if desired, and the results are in agreement with the various partial results of the synthetic solution in Love's 'Mathematical Theory of Elasticity,' chapter xxii.

9. *The Displacements in Terms of Complex Potentials.*

From (7.1)

$$8\mu\mathrm{D}=f(z, \mathrm{Z})-\{z\bar{\Omega}_0'(\bar{z})+\bar{\omega}_0'(\bar{z})\}+(1-\eta)\left\{z\bar{\Omega}_1'(\bar{z})+\bar{\omega}_1'(\bar{z})+\frac{3p}{8h^3}z^2\bar{z}\right\}\mathrm{Z}$$
$$+2\sigma\bar{\Omega}_0''(\bar{z})\mathrm{Z}^2-\tfrac{2}{3}(2-\eta)\bar{\Omega}_1''(\bar{z})\mathrm{Z}^3. \quad . \quad (9.1)$$

Comparison of the expressions for $\partial\mathrm{D}/\partial z+\partial\bar{\mathrm{D}}/\partial\bar{z}$ from (9.1) and (7.3), using (8.1) and (8.3)–(8.7), leads to

$$f(z, \mathrm{Z})=(3-4\sigma)\Omega_0(z)+2\sigma pz-(2-\eta)\frac{pz}{2h^3}\mathrm{Z}^3$$
$$+\left\{(1-\eta)\Omega_1(z)+3p\sigma\frac{z}{h}\right\}\mathrm{Z}+iz\mathrm{F}(\mathrm{Z}), \quad . \quad (9.2)$$

where F(Z) is real. Hence

$$\mathrm{D}=\mathrm{D}_0+\mathrm{D}_1\mathrm{Z}+\mathrm{D}_2\mathrm{Z}^2+\mathrm{D}_3\mathrm{Z}^3+iz\mathrm{F}(\mathrm{Z}), \quad . \quad . \quad . \quad . \quad . \quad . \quad (9.3)$$

where

$$8\mu\mathrm{D}_0=(3-4\sigma)\Omega_0(z)-z\bar{\Omega}_0'(\bar{z})-\bar{\omega}_0'(\bar{z})+2\sigma pz, \quad . \quad . \quad . \quad . \quad . \quad (9.4)$$

$$8\mu\mathrm{D}_1=(1-\eta)\left\{\Omega_1(z)+z\bar{\Omega}_1'(\bar{z})+\bar{\omega}_1'(\bar{z})+\frac{3p}{8h^3}z^2\bar{z}\right\}+3\sigma\frac{pz}{h}. \quad . \quad (9.5)$$

$$8\mu\mathrm{D}_2=2\sigma\bar{\Omega}_0''(\bar{z}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9.6)$$

$$8\mu\mathrm{D}_3=-(2-\eta)\left\{\frac{2}{3}\bar{\Omega}_1''(\bar{z})+\frac{pz}{2h^3}\right\}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9.7)$$

Again, from (7.4) and (8.1)–(8.7), we find

$$w=w_0+w_1\mathrm{Z}+w_2\mathrm{Z}^2+w_4\mathrm{Z}^4, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9.8)$$

where $$2\mu w_0 = f(z, \bar{z}),$$

$$2\mu w_1 = -\tfrac{1}{2}\sigma\{\Omega_0'(z)+\bar{\Omega}_0'(\bar{z})\}-\tfrac{1}{2}p(1-\sigma), \quad \ldots \ldots \quad (9.9)$$

$$2\mu w_2 = -\tfrac{1}{4}\eta\left\{\Omega_1'(z)+\bar{\Omega}_1'(\bar{z})+\frac{3}{4}\frac{p}{h^3}z\bar{z}\right\}-\frac{3p}{8h}(1-\sigma), \quad \ldots \quad (9.10)$$

and $$2\mu w_4 = (1+\eta)\frac{p}{16h^3}. \quad \ldots \ldots \quad (9.11)$$

These make

$$2\mu\frac{\partial w}{\partial\bar{z}} = \frac{\partial f}{\partial\bar{z}}(z, \bar{z})-\tfrac{1}{2}\sigma\bar{\Omega}_0''(\bar{z})\mathrm{Z}-\tfrac{1}{4}\eta\left\{\bar{\Omega}_1''(\bar{z})+\frac{3pz}{4h^3}\right\}\mathrm{Z}^2,$$

whereas from (7.2), (8.8) and (9.3)

$$2\mu\frac{\partial w}{\partial\bar{z}} = \frac{3pz}{8h}(1-\sigma)+\tfrac{1}{2}h^2\bar{\Omega}_1''(\bar{z})-\tfrac{1}{8}(1-\eta)\left\{\Omega_1(z)+z\bar{\Omega}_1'(\bar{z})+\bar{\omega}_1'(\bar{z})+\frac{3p}{8h^3}z^2\bar{z}\right\}$$
$$-\tfrac{1}{2}\sigma\bar{\Omega}_0''(\bar{z})\mathrm{Z}-\tfrac{1}{4}\eta\left\{\bar{\Omega}_1''(\bar{z})+\frac{3pz}{4h^3}\right\}\mathrm{Z}^2-i\mu z\mathrm{F}'(\mathrm{Z}).$$

Comparing these two expressions for $\partial w/\partial\bar{z}$, we deduce that $\mathrm{F}'(\mathrm{Z})$ must be a real constant, but we can drop the corresponding terms from w and D, as they merely represent a rigid body displacement. We then find $f(z, \bar{z})$, and remembering that w_0 must be real, we have

$$2\mu w_0 = \tfrac{1}{2}h^2\left\{\Omega_1'(z)+\bar{\Omega}_1'(\bar{z})+\tfrac{3}{4}(1-\sigma)\frac{p}{h^3}z\bar{z}\right\}$$
$$-\tfrac{1}{8}(1-\eta)\left\{\bar{z}\Omega_1(z)+z\bar{\Omega}_1(\bar{z})+\omega_1(z)+\bar{\omega}_1(\bar{z})+\frac{3}{16}\frac{p}{h^3}z^2\bar{z}^2\right\}. \quad (9.12)$$

10. *The Stress and Couple Resultants in Terms of Complex Potentials.*

From their definitions in (3.1)–(3.5) we find, using (8.4)–(8.11),

$$\Theta_0 = \theta_0 = \tfrac{1}{2}\{\Omega_0'(z)+\bar{\Omega}_0'(\bar{z})\}, \quad \ldots \ldots \quad (10.1)$$

$$\Phi_0 = \phi_0+\tfrac{1}{3}h^2\phi_2 = -\tfrac{1}{2}\{z\bar{\Omega}_0''(\bar{z})+\bar{\omega}_0''(\bar{z})+\tfrac{2}{3}h^3\sigma\bar{\Omega}_0'''(\bar{z})\}, \quad \ldots \quad (10.2)$$

$$\Psi_0 = -\tfrac{2}{3}h^2\psi = \tfrac{1}{3}h^2\left\{\bar{\Omega}_1''(\bar{z})+\frac{3pz}{4h^3}\right\}, \quad \ldots \ldots \quad (10.3)$$

$$\Lambda = \tfrac{1}{3}h^2\theta_1+\tfrac{1}{5}h^4\theta_3, \quad \Gamma = \tfrac{1}{3}h^2\phi_1+\tfrac{1}{5}h^4\phi_3,$$

or $$\Lambda = \tfrac{1}{6}h^2(1+\eta)\left\{\Omega_1'(z)+\bar{\Omega}_1'(\bar{z})+\frac{3p}{4h^3}z\bar{z}\right\}-(2+\eta)\frac{ph}{20}, \quad \ldots \quad (10.4)$$

$$\Gamma = \tfrac{1}{6}h^2(1-\eta)\left\{z\bar{\Omega}_1''(\bar{z})+\bar{\omega}_1''(\bar{z})+\frac{3pz^2}{8h^3}\right\}-\frac{1}{15}(2-\eta)h^4\bar{\Omega}_1'''(z). \quad (10.5)$$

11. *Curvature of the Mid-Plane in Terms of the Complex Potentials.*

From (5.1), (5.2) and (9.12), we have

$$2\mu\beta = 2h^2\bar{\Omega}_1'''(\bar{z})-\tfrac{1}{2}(1-\eta)\left\{z\bar{\Omega}_1''(\bar{z})+\bar{\omega}_1''(\bar{z})+\frac{3pz^2}{8h^3}\right\}. \quad \ldots \quad (11.1)$$

$$2\mu\gamma=\frac{3p}{2h}(1-\sigma)-\tfrac{1}{2}(1-\eta)\left\{\Omega_1{}'(z)+\bar{\Omega}_1{}'(\bar{z})+\frac{3pz\bar{z}}{4h^3}\right\}. \quad . \quad . \quad (11.2)$$

Hence, also, from (10.4)–(11.2), we have

$$\Gamma=-\mathrm{P}(1-\eta)\beta+(8+\eta)\frac{h^4}{15}\bar{\Omega}_1{}'''(\bar{z}), \quad . \quad . \quad . \quad . \quad (11.3)$$

$$\Lambda=-\mathrm{P}(1+\eta)\gamma+\frac{(8+\eta+\eta^2)ph}{20(1-\eta)}, \quad . \quad . \quad . \quad . \quad . \quad (11.4)$$

where P is a plate modulus or flexural rigidity of the plate defined by

$$\mathrm{P}=2h^2\mu/3(1-\eta). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11.5)$$

Note that it is only in the cases where p and $\Omega_1{}'''(z)$ vanish together that we have the results (6.4), which are practically the starting point of the elementary theory. It will be seen from the particular examples treated later that the thinner the plate, the less important are these correcting terms. When $p=0$, $\widehat{zz}=0$; also when $\Omega_1{}'''(z)=0$ in addition, it is clear from (11.2) that the sum of the principal curvatures is linear in x and y. Hence the elementary results (6.4) for the stress couples are exact when the plate is in a state of generalized plane stress with the mid-plane bent so that the sum of the principal curvatures is linear in x and y (*cp.* Love's 'Mathematical Theory of Elasticity,' p. 474).

It should be remarked that the mid-plane deflexion w_0 satisfies (6.8) exactly.

12. *Boundary Conditions.*

Broadly speaking, the complex potentials $\Omega_0(z)$, $\omega_0(z)$ are concerned with the stretching of the plate in its own plane, since $\widehat{nn}$, $\widehat{ns}$, D_0 involve only $\Omega_0(z)$, $\omega_0(z)$. Again, $\Omega_1(z)$, $\omega_1(z)$ are intimately concerned with the bending of the plate out of its equilibrium plane, since $\widetilde{nz}$, $\widetilde{ns}$, $\widetilde{nn}$, w_0 involve only $\Omega_1(z)$, $\omega_1(z)$.

The boundary conditions may be imposed in terms of displacements, or in terms of displacements and stress couples. Thus the semi-inverse solution is wide enough to deal with a "horizontally" clamped or built-in edge defined by

$$w_0=0, \quad \frac{\partial w_0}{\partial n}=0, \quad \mathrm{D}_0=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (12.1)$$

Here the solution for w_0 is the same as in the elementary theory, but the solutions for the stress couple combinations Λ, Γ may differ slightly. The clamping couple $\widetilde{ns}$ is found as

$$4\widetilde{ns}=\Gamma'+\bar{\Gamma}'+2\Lambda, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12.2)$$

and in general varies round the boundary. For a supported boundary under a constant flexural couple M, we have to satisfy

$$w_0=0, \quad \widetilde{ns}=\mathrm{M}, \quad \mathrm{D}_0=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12.3)$$

round the boundary, and here the elementary and exact semi-inverse solutions will generally differ by terms which become less important the thinner the plate. An important case is the *simply supported* boundary where M=0.

As regards the conditions for D_0 in the above boundary relations, these can always be satisfied by making the mid-plane displacement D_0 vanish everywhere. This can be done by supplying an all-round constant pressure resultant round the boundary. For if we take

$$\Omega_0(z)=-\frac{\eta}{1-\eta}pz, \quad \omega_0(z)=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (12.4)$$

then, from (9.4), D_0 vanishes everywhere, whilst from (3.8), (10.1) and (10.2) we have round any boundary

$$\Theta_0'=\Theta_0=-\frac{\eta}{1-\eta}p, \quad \Phi_0'=e^{-2i\alpha}\Phi_0=0,$$

whence

$$\breve{ns}=0, \quad \breve{nn}=-\frac{1}{2}\frac{\eta}{1-\eta}p, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12.5)$$

giving the amount of the all-round pressure as $\eta p/2(1-\eta)$.

The boundary problems (12.1) and (12.3) involve definite distributions of the " vertical " shear $\breve{nz}$ and the torsional stress couple $\widetilde{nn}$ in thick plates. In moderately thick plates we can invoke Saint-Venant's principle to allow us to consider the solution as that corresponding to any equivalent distribution of boundary shear and torsional stress couple except for points on and near to the bounding edge of the plate. Thus, in particular, since the distribution of differentiable stress couple $\widetilde{nn}$ is statically equivalent to a distribution of shear resultant $\breve{nz}$ of amount $-\partial\widetilde{nn}/\partial s$, we might replace all the torsional stress couple $\widetilde{nn}$ by vertical shear (*cp.* Love's 'Mathematical Theory of Elasticity,' p. 460). The semi-inverse plate solutions have therefore the same type of validity at the bounding edges as the Saint-Venant semi-inverse solutions for torsion and flexure with shear have at the ends of the beam.

If, in the replacement of $\widetilde{nn}$ by vertical shear $\breve{nz}$, $\widetilde{nn}$ is discontinuous at a point P on the mid-plane boundary curve C, so that on passing through P, keeping the area enclosed on the left, $\widetilde{nn}$ increases suddenly by $\Delta\widetilde{nn}$, then in the replacement we have an isolated force R at P given by

$$R=-2h\Delta\widetilde{nn}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12.6)$$

13. *The Equilateral Triangular Plate.*

Let the origin be the centroid of an equilateral triangle of altitude $3a$, and let one edge given by $z+\bar{z}=2a$. Since rotation to new axes, given by $z'=ze^{\pm 2i\pi/3}$ gives the remaining edges in turn, the composite equation of the boundary in complex co-ordinates is

$$(z+\bar{z}-2a)(ze^{2i\pi/3}+\bar{z}e^{-2i\pi/3}-2a)(ze^{-2i\pi/3}+\bar{z}e^{2i\pi/3}-2a)=0$$

or

$$z^3+\bar{z}^3+6az\bar{z}-8a^3=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (13.1)$$

Assume the tentative solution for w_0, where α and β are real constants,

$$2\mu w_0=(z^3+\bar{z}^3+6az\bar{z}-8a^3)(\alpha z\bar{z}+\beta) \quad . \quad . \quad . \quad . \quad (13.2)$$

or

$$2\mu w_0=6a\alpha z^2\bar{z}^2+\alpha z\bar{z}(z^3+\bar{z}^3)+(6a\beta-8a^3\alpha)z\bar{z}+\beta(z^3+\bar{z}^3)-8a^3\beta, \quad (13.3)$$

which vanishes round the boundary (13.1). Notice also that the rotations leave the form of (13.2) unaltered, so that the solution has the symmetry appropriate to equal distributions along the three sides of the triangle.

Now, from (9.12), the complex potentials

$$\Omega_1(z)=\mathrm{A}z^4+\mathrm{B}z, \quad \omega_1(z)=\mathrm{C}z^3+\mathrm{E}, \quad . \quad . \quad . \quad . \quad (13.4)$$

where A, B, C, E are real constants, lead to

$$2\mu w_0=2\mathrm{A}h^2(z^3+\bar{z}^3)+\mathrm{B}h^2+\tfrac{3}{8}(1-\sigma)\frac{p}{h}z\bar{z}$$

$$-\tfrac{1}{8}(1-\eta)\left\{\mathrm{A}z\bar{z}(z^3+\bar{z}^3)+2\mathrm{B}z\bar{z}+\mathrm{C}(z^3+\bar{z}^3)+2\mathrm{E}+\frac{3p}{16h^3}z^2\bar{z}^2\right\}. \quad (13.5)$$

Comparison of (13.3) and (13.5) gives

$$\alpha=-\frac{(1-\eta)p}{256ah^3}=-\frac{\mu p}{384ah\mathrm{P}}, \quad \mathrm{A}=\frac{p}{32ah^3}, \quad . \quad . \quad (13.6)$$

$$\tfrac{1}{4}(1-\eta)\mathrm{B}=\tfrac{3}{8}(1-\sigma)\frac{p}{h}-(1-\eta)\frac{pa^2}{32h^3}-6a\beta, \quad . \quad . \quad . \quad . \quad . \quad (13.7)$$

$$\tfrac{1}{8}(1-\eta)\mathrm{C}=\frac{p}{16ah}-\beta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13.8)$$

$$\tfrac{1}{4}(1-\eta)\mathrm{E}=h^2\mathrm{B}+8a^3\beta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13.9)$$

From (10.4), using (13.6) and (13.1), Λ is constant along the boundary and equal there to

$$\Lambda=\tfrac{1}{3}h^2(1+\eta)\left(\mathrm{B}+\frac{pa^2}{2h^3}\right)-(2+\eta)\frac{ph}{20}. \quad . \quad . \quad . \quad (13.10)$$

From (10.5), using (13.6), we have

$$\Gamma+\bar{\Gamma}=h^2(1-\eta)\left\{\frac{pz\bar{z}}{16ah^3}(z+\bar{z}-2a)+\mathrm{C}(z+\bar{z})+\frac{p}{16h^3}(z+\bar{z})^2\right\}$$

$$-(2-\eta)\frac{ph}{20a}(z+\bar{z}),$$

and along the side $z+\bar{z}=2a$ this has the constant value

$$\Gamma+\bar{\Gamma}=h^2(1-\eta)\left\{2a\mathrm{C}+\frac{pa^2}{4h^3}\right\}-(2+\eta)\frac{ph}{10}. \quad . \quad . \quad . \quad (13.11)$$

Hence, for a constant flexural couple $\widetilde{xy}=\mathrm{M}$ along this side, and by symmetry all round the boundary,

$$4\mathrm{M}=\Gamma+\bar{\Gamma}+2\Lambda=(1-\eta)\left\{2a\mathrm{C}h^2+\frac{pa^2}{4h}\right\}+\tfrac{1}{3}(1+\eta)\left\{2\mathrm{B}h^2+\frac{pa^2}{h}\right\}-\tfrac{2}{5}ph.$$

$$. \quad . \quad . \quad (13.12)$$

From (13.6)–(13.8), using (7.5), (13.12) gives

$$-\frac{\beta}{\alpha}=\frac{8}{5}\,\frac{8-3\eta}{1-\eta}h^2+4a^2-32\mathrm{M}\frac{h}{p}, \quad . \quad . \quad . \quad . \quad (13.13)$$

whence from (13.2) and (11.5)

$$w_0=\frac{p}{384ah\mathrm{P}}\{3xy^2-x^3-3a(x^2+y^2)+4a^3\}\left(x^2+y^2+\frac{\beta}{\alpha}\right). \quad (13.14)$$

This extends the result due to Woinowsky–Krieger as given in Timoshenko's 'Theory of Plates and Shells' (1940, p. 293) for the simply supported case $M=0$ with the approximations of the elementary theory, corresponding in our notation to $-\beta/\alpha=4a^2$.

From (13.7)–(13.9), and (13.13) we have for the remaining constants B, C, E of the complex potentials

$$B=\frac{3p}{20h}\,\frac{2-3\eta}{1+\eta}-\frac{pa^2}{2h^3}+\frac{3M}{h^2}, \qquad (13.15)$$

$$C=\frac{p}{20ah}\,\frac{2+3\eta}{1-\eta}-\frac{pa}{8h^3}+\frac{M}{ah^2}, \qquad (13.16)$$

$$E=\frac{3}{5}ph\,\frac{2-3\eta}{1-\eta^2}-\frac{1}{5}\,\frac{pa^2}{h}\,\frac{2+3\eta}{1-\eta}+\frac{pa^4}{2h^3}+4M\left\{\frac{3}{1-\eta}-\frac{a^2}{h^2}\right\}. \qquad (13.17)$$

If we set $p=0$, we have the case of the equilateral triangular plate under constant flexural edge couple M given by the complex potentials

$$\Omega_1(z)=3\frac{M}{h^2}z, \quad \omega_1(z)=\frac{Mz^3}{ah^2}+4M\left\{\frac{3}{1-\eta}-\frac{a^2}{h^2}\right\}. \qquad (13.18)$$

For the deflexion at the centroid we find $Ma^2/3P$ in agreement with the result given by Timoshenko (p. 103). Clearly the elementary solution is also here the exact semi-inverse solution, since in (11.3) and (11.4) both p and $\Omega_1'''(z)$ are zero.

Returning to our general solution, we remark that the solution corresponds to a torsional couple $\widetilde{xx}=ky^3+k'y$, where k and k' are constants, and to a parabolic shear resultant

$$\widetilde{nz}=\widetilde{xz}=\frac{p}{8ah}(3a^2-y^2) \qquad (13.19)$$

along $x=a$. If a is considerably larger than h, the solution can also, by Saint-Venant's principle, be taken to be that for which there is no torsional couple along the edges, that given by the solution being re-distributed as the equivalent shear $-\partial\widetilde{xx}/\partial y$ along $x=a$, and similarly for the other sides (thus giving a different, but still parabolic distribution of shear), together with isolated forces at the vertices of the triangle.

14. *The Circular Plate.*

If $z=re^{i\theta}$, the solution of problems symmetrical in r is given by taking the complex potentials

$$\Omega_1(z)=Az+Ez\log z, \quad \omega_1(z)=B+F\log z, \qquad (14.1)$$

where the constants A, B, E, F are real. For a complete circular plate we put E and F zero to avoid infinite stresses and displacements at the origin, and (9.12) and (14.1) then give

$$2\mu w_0=h^2A-\tfrac{1}{4}(1-\eta)B+r^2\left\{\frac{3p}{8h}(1-\sigma)-\tfrac{1}{4}(1-\eta)A\right\}-\frac{3}{128}\,\frac{p}{h^3}(1-\eta)r^4, \qquad (14.2)$$

which we can write as

$$w_0 = -\frac{p}{128h\mathrm{P}}(r^2-a^2)^2 + \mathrm{C}(r^2-a^2), \quad . \quad . \quad . \quad . \quad (14.3)$$

vanishing round the circular boundary $r=a$, if we take

$$2\mu\mathrm{C} = \tfrac{1}{4}(1-\eta)\frac{\mathrm{B}}{a^2} - \frac{h^2}{a^2}\mathrm{A} - \frac{3}{128}\frac{p}{h^3}(1-\eta)a^2, \quad . \quad . \quad . \quad (14.4)$$

and

$$2\mu\mathrm{C} = \frac{3p}{8h}(1-\sigma) - \tfrac{1}{4}(1-\eta)\mathrm{A} - \frac{3}{64}\frac{p}{h^3}(1-\eta)a^2. \quad . \quad . \quad (14.5)$$

Also (3.8), (10.3)–(10.5) give

$$\mathit{\Lambda}' = \widetilde{r\theta} - \widetilde{\theta r} = \tfrac{1}{6}h^2(1+\eta)\left\{2\mathrm{A} + \frac{3}{4}\frac{pr^2}{h^3}\right\} - (2+\eta)\frac{ph}{20}, \quad \mathit{\Psi}_0' = \breve{rz} + i\breve{\theta z} = \frac{pr}{4h}, \quad . \quad . \quad . \quad (14.6)$$

$$\mathit{\Gamma}' = \widetilde{r\theta} + \widetilde{\theta r} + 2i\widetilde{\theta\theta} = e^{-2i\theta}\mathit{\Gamma} = \frac{1}{16}\frac{p}{h}(1-\eta)r^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14.7)$$

and the solution gives a constant flexural couple M given by

$$\mathrm{M} = \widetilde{r\theta} = \tfrac{1}{4}\{\mathit{\Gamma}' + \bar{\mathit{\Gamma}}' + 2\mathit{\Lambda}'\},$$

round the boundary $r=a$, if we take

$$\mathrm{A} = \frac{6}{(1+\eta)h^2}\left\{\mathrm{M} - \frac{1}{32}\frac{p}{h}(3+\eta)a^2 + (2+\eta)\frac{ph}{40}\right\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14.8)$$

$$\mathrm{B} = \frac{6\mathrm{M}}{(1+\eta)h^2}\left\{\frac{4h^2}{1-\eta} - a^2\right\} + \frac{3}{5}\frac{2+\eta}{1-\eta^2}ph + \frac{3}{20}\frac{pa^2}{h}\frac{\eta^2-4\eta-7}{1-\eta^2} + \frac{3}{32}\frac{pa^4}{h^3}\frac{5+\eta}{1+\eta}, \quad . \quad . \quad . \quad (14.9)$$

$$\mathrm{C} = \frac{1}{2(1+\eta)\mathrm{P}}\left\{\frac{pa^2}{16h} + \frac{ph}{40}\frac{8+\eta+\eta^2}{1-\eta} - \mathrm{M}\right\}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14.10)$$

The solution makes $\widetilde{\theta\theta}=0$ everywhere and $\breve{rz}=pa/4h$ round the boundary, as could have been deduced from considerations of symmetry and the external forces on the plate. The clamping couple M_0 and the deflexion w_0 corresponding to the built-in or " horizontally " clamped edge, for which $(\partial w_0/\partial r)_{r=a}=0$, are given by $\mathrm{C}=0$, or

$$\mathrm{M}_0 = \frac{1}{16}\frac{pa^2}{h} + \frac{ph}{40}\frac{8+\eta+\eta^2}{1-\eta}, \quad w_0 = -\frac{p}{128h\mathrm{P}}(r^2-a^2)^2. \qquad (14.11)$$

For the simply supported plate $\mathrm{M}=0$, and in this case

$$w_0 = -\frac{p}{128h\mathrm{P}}(r^2-a^2)^2 + \frac{\mathrm{M}_0}{2(1+\eta)\mathrm{P}}(r^2-a^2), \quad . \quad . \quad (14.12)$$

where M_0 is given by (14.11) ; the central deflexion $w_0=\mathrm{W}$ is given by

$$2\mu\mathrm{W} = -\frac{3}{80}\frac{pa^2}{h}\frac{8+\eta+\eta^2}{1+\eta} - \frac{3}{128}\frac{pa^4}{h^3}\left(\frac{1-\eta}{1+\eta}\right)(5+\eta). \quad . \quad (14.13)$$

For the circular plate under a constant flexural edge couple M only, we put $p=0$ and the solutions for w_0 and W are

$$w_0 = \frac{\mathrm{M}}{2(1+\eta)\mathrm{P}}(a^2-r^2), \quad 2\mu\mathrm{W} = \frac{3a^2}{2h^2}\left(\frac{1-\eta}{1+\eta}\right)\mathrm{M}, \quad . \quad . \quad (14.14)$$

the corresponding complex potentials being

$$\Omega_1(z)=\frac{6\mathrm{M}z}{(1+\eta)h^2},\quad \omega_1(z)=\frac{6\mathrm{M}}{(1+\eta)h^2}\left\{\frac{4h^2}{1-\eta}-a^2\right\}. \quad . \quad (14.15)$$

These well-known results, given anew in terms of complex potentials, will be needed for comparison with corresponding results for the elliptic plate. The corresponding problems for the circular annulus present no difficulty and we proceed to the elliptic plate problems, where the text-books give only that for the clamped elliptic plate, due to Bryan, which happens to retain the same algebraic simplicity as in the circular case and corresponds to simple polynomial complex potentials in the present mode of treatment.

15. *The Elliptic Plate.*

We take the elliptic boundary as given by $\xi=\alpha$ in the orthogonal curvilinear network

$$z=c\cosh\zeta,\quad \zeta=\xi+i\eta \quad . \quad . \quad . \quad . \quad . \quad . \quad (15.1)$$

(as we shall not need η explicitly, there will be no occasion to confuse it with Poisson's ratio in what follows). The semi-axes of the boundary are given by

$$a=c\cosh\alpha,\quad b=c\sinh\alpha,\quad c^2=a^2-b^2 \quad . \quad . \quad . \quad . \quad (15.2)$$

It is clear that

$$w_0=\mathrm{W}\left(1-\frac{x^2}{a^2}-\frac{y^2}{b^2}\right)^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (15.3)$$

satisfies $\nabla_1^4 w_0=-p/2h\mathrm{P}$, $\partial w_0/\partial n=0$ round the boundary if

$$\mathrm{W}=-a^4b^4p/16h\mathrm{P}(3a^4+3b^4+2a^2b^2), \quad . \quad . \quad . \quad . \quad (15.4)$$

which is Bryan's result (*cp.* Love's 'Mathematical Theory of Elasticity,' p.484). To find the complex potentials in this case, since

$$4a^2b^2\left(1-\frac{x^2}{a^2}-\frac{y^2}{b^2}\right)=4a^2b^2+(a^2-b^2)(z^2+\bar{z}^2)-2(a^2+b^2)z\bar{z},$$

we have

$$\begin{aligned}2\mu w_0=\frac{\mu\mathrm{W}}{8a^4b^4}\{&16a^4b^4+8a^2b^2(a^2-b^2)(z^2+\bar{z}^2)-16a^2b^2(a^2+b^2)z\bar{z}\\&+(a^2-b^2)^2(z^4+\bar{z}^4)-4(a^4-b^4)z\bar{z}(z^2+\bar{z}^2)\\&+2(3a^4+3b^4+2a^2b^2)z^2\bar{z}^2\}. \quad . \quad (15.5)\end{aligned}$$

Taking the complex potentials, where A, B, C, E, F are real constants,

$$\Omega_1(z)=\mathrm{A}z^3+\mathrm{B}z,\quad \omega_1(z)=\mathrm{C}z^4+\mathrm{E}z^2+\mathrm{F}, \quad . \quad . \quad . \quad (15.6)$$

we have from (9.12)

$$\begin{aligned}2\mu w_0=h^2&\left\{\mathrm{B}+\tfrac{3}{2}\mathrm{A}(z^2+\bar{z}^2)+\tfrac{3}{8}(1-\sigma)\frac{p}{h^3}z\bar{z}\right\}\\&-\tfrac{1}{8}(1-\eta)\left\{\mathrm{A}z\bar{z}(z^2+\bar{z}^2)+2\mathrm{B}z\bar{z}+\mathrm{C}(z^4+\bar{z}^4)+\mathrm{E}(z^2+\bar{z}^2)+2\mathrm{F}+\frac{3}{16}\frac{p}{h^3}z^2\bar{z}^2\right\}.\\&\quad . \quad . \quad . \quad (15.7)\end{aligned}$$

Comparing coefficients in (15.5) and (15.7), we have, in addition to (15.4), the five constants

$$\mathrm{A}=\frac{4\mu\mathrm{W}(a^4-b^4)}{(1-\eta)a^4b^4}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15.8)$$

$$\mathrm{B}=\frac{3p}{2h}\,\frac{1}{1-\eta^2}+\frac{8\mu\mathrm{W}(a^2+b^2)}{(1-\eta)a^2b^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15.9)$$

$$\mathrm{C}=-\frac{\mu\mathrm{W}(a^2-b^2)^2}{(1-\eta)a^4b^4}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15.10)$$

$$\mathrm{E}=\frac{48\mu\mathrm{W}h^2(a^4-b^4)}{(1-\eta)^2a^4b^4}-\frac{8\mu\mathrm{W}(a^2-b^2)}{(1-\eta)a^2b^2}, \quad . \quad . \quad . \quad . \quad . \quad (15.11)$$

$$\mathrm{F}=\frac{6ph}{(1-\eta)^2(1+\eta)}+\frac{32\mu h^2\mathrm{W}(a^2+b^2)}{a^2b^2(1-\eta)^2}-\frac{8\mu\mathrm{W}}{1-\eta}. \quad . \quad . \quad (15.12)$$

If $\widetilde{\xi\eta}-\widetilde{\eta\xi}=\Lambda'$, $\widetilde{\xi\eta}+\widetilde{\eta\xi}+2i\widetilde{\eta\eta}=\Gamma'$, $\widetilde{\xi\eta}=\frac{1}{4}(\Gamma'+\bar{\Gamma}'+2\Lambda')$, (15.13)

the equations of transformation of the stress couples are, from (1.10), (3.8), and (15.1), given by

$$\Lambda'=\Lambda, \quad \Gamma'=\frac{\sinh\bar{\zeta}}{\sinh\zeta}\Gamma=\frac{\varpi^2}{a^2b^2}(\bar{z}^2-c^2)\Gamma, \quad . \quad . \quad . \quad (15.14)$$

where ϖ is the perpendicular from the origin on the tangent at $z=\zeta$ to the ellipse $\xi=$constant, passing through the point $z=\zeta$. We shall find the flexural clamping couple $\widetilde{\xi\eta}$ in terms of the perpendicular on $\xi=\alpha$. For the remaining elliptic plate problems contemplated, we need further complex potentials calculated to annul this couple or reduce it to a constant edge couple. This is found to need careful attention to the admissibility of the stresses at the foci $z=\pm c$ or $\zeta=0$, $\zeta=i\pi$, where $dz/d\zeta$ vanishes, and the complex potentials which give the required type of boundary stress couples have to be so chosen that their combination gives rise to no infinite stress couples at the foci. The complex potentials already considered being simple polynomials in z, are readily seen to lead to admissible stresses, couples and displacements throughout the plate.

It is not difficult to show that along the boundary $\zeta+\bar{\zeta}=2\alpha$, we have

$$z\bar{z}=c^2\cosh 2\alpha-\frac{a^2b^2}{\varpi^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (15.15)$$

$$z^2+\bar{z}^2-c^2=\cosh 2\alpha\left\{c^2\cosh 2\alpha-\frac{2a^2b^2}{\varpi^2}\right\}, \quad . \quad (15.16)$$

$$c^2\{\sinh 2\zeta+\sinh 2\bar{\zeta}\}=2\sinh 2\alpha\left\{c^2\cosh 2\alpha-\frac{2a^2b^2}{\varpi^2}\right\}. \quad . \quad (15.17)$$

We next find the displacements and stress couples along the boundary corresponding to the separate complex potentials. Thus from

$$\Omega_1(z)=\mathrm{B}z, \quad \omega_1(z)=\mathrm{F}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15.18)$$

(9.12) gives, using (15.15), a boundary displacement

$$2\mu w_0 = \frac{3}{8}(1-\sigma)\frac{p}{h}\left(c^2\cosh 2\alpha - \frac{a^2b^2}{\varpi^2}\right) - \tfrac{1}{4}(1-\eta)\mathrm{F}$$
$$+\mathrm{B}h^2 - \tfrac{1}{4}(1-\eta)\mathrm{B}\left(c^2\cosh 2\alpha - \frac{a^2b^2}{\varpi^2}\right)$$
$$-\frac{3}{128}(1-\eta)\frac{p}{h^3}\left\{c^4\cosh^2 2\alpha - \frac{2a^2b^2}{\varpi^2}c^2\cosh 2\alpha + \frac{a^4b^4}{\varpi^4}\right\}. \quad (15.19)$$

The deflexion W at $z=0$ is given by

$$2\mu\mathrm{W} = \mathrm{B}h^2 - \tfrac{1}{4}(1-\eta)\mathrm{F}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15.20)$$

Again from (10.4) and (10.5), we have

$$\Lambda = (1+\eta)\left\{\frac{1}{3}h^2\mathrm{B} + \frac{1}{8}\frac{p}{h}z\bar{z}\right\} - (2+\eta)\frac{ph}{20}, \quad \Gamma = \frac{1}{16}(1-\eta)\frac{p}{h}z^2,$$

leading, from (15.14)–(15.16), to a boundary flexural couple $\widetilde{\xi\eta}$, given by

$$4\widetilde{\xi\eta} = \tfrac{1}{8}(1-\eta)\frac{p}{h}\left\{2\varpi^2 - c^2\cosh 2\alpha + \frac{a^2b^2}{\varpi^2}\right\} + \tfrac{2}{3}h^2(1+\eta)\mathrm{B}$$
$$+\tfrac{1}{4}(1+\eta)\frac{p}{h}\left\{c^2\cosh 2\alpha - \frac{a^2b^2}{\varpi^2}\right\} - (2+\eta)\frac{ph}{10}. \quad . \quad (15.21)$$

The remaining complex potentials are now of a complementary nature, *i. e.*, in finding their corresponding displacements, stresses and couples, we must put $p=0$. In each case the central deflexion W at $z=0$ or $\zeta = \pm i\pi/2$ will be found to vanish. Proceeding as before, we find from the complex potentials

$$\Omega_1(z) = 0, \quad \omega_1(z) = \mathrm{C}z^4 + \mathrm{E}z^2, \quad . \quad . \quad . \quad . \quad . \quad (15.22)$$

that the boundary displacement is given by

$$2\mu w_0 = -\tfrac{1}{8}(1-\eta)\mathrm{C}\left\{c^4(1+\cosh^4 2\alpha) - \frac{4a^2b^2}{\varpi^2}c^2\cosh^3 2\alpha + \frac{2a^4b^4}{\varpi^4}\cosh 4x\right\}$$
$$-\tfrac{1}{8}(1-\eta)\mathrm{E}\left\{c^2(1+\cosh^2 2\alpha) - \frac{2a^2b^2}{\varpi^2}\cosh 2\alpha\right\}, \quad . \quad . \quad (15.23)$$

and the stress couples by

$$\Lambda = 0, \quad \Gamma = 2h^2(1-\eta)\mathrm{C}\bar{z}^2 + \tfrac{1}{3}(1-\eta)h^2\mathrm{E}$$

leading, from (15.14)–(15.16), to a boundary flexural couple given by

$$4\widetilde{\xi\eta} = 4h^2(1-\eta)\mathrm{C}\left\{2\varpi^2\cosh^2 2\alpha - c^2\cosh 2\alpha\cosh 4\alpha + \frac{a^2b^2}{\varpi^2}\cosh 4\alpha\right\}$$
$$+\tfrac{2}{3}(1-\eta)\mathrm{E}\frac{h^2}{c^2}\{2\varpi^2 - c^2\cosh 2\alpha\}. \quad (15.24)$$

Similarly, for

$$\Omega_1(z) = \mathrm{A}z^3, \quad \omega_1(z) = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (15.25)$$

the boundary displacement can be written

$$2\mu w_0=\tfrac{3}{2}h^2\mathrm{A}\left\{c^2(1+\cosh^2 2\alpha)-\frac{2a^2b^2}{\varpi^2}\cosh 2\alpha\right\}$$
$$-\tfrac{1}{8}(1-\eta)\mathrm{A}\left\{c^4(\cosh 2\alpha+\cosh^3 2\alpha)-\frac{a^2b^2}{\varpi^2}c^2(1+3\cosh^2 2\alpha)+\frac{2a^4b^4}{\varpi^4}\cosh 2\alpha\right\}. \quad . \quad (15.26)$$

The stress couples are given by

$$\Lambda=\tfrac{1}{2}h^2(1+\eta)\mathrm{A}(z^2+\bar{z}^2),\quad \Gamma=h^2(1-\eta)\mathrm{A}z\bar{z}-\tfrac{2}{5}(2-\eta)\mathrm{A}h^4$$

and the boundary flexural couple $\widetilde{\xi\eta}$ by

$$4\widetilde{\xi\eta}=\mathrm{A}h^2(1-\eta)\left\{4\varpi^2\cosh 2\alpha-c^2(3\cosh^2 2\alpha-1)+\frac{2a^2b^2}{\varpi^2}\cosh 2\alpha\right\}$$
$$+\mathrm{A}h^2(1+\eta)\left\{c^2(1+\cosh^2 2\alpha)-\frac{2a^2b^2}{\varpi^2}\cosh 2\alpha\right\}$$
$$-\tfrac{4}{5}\mathrm{A}\frac{h^4}{c^2}(2-\eta)\{2\varpi^2-c^2\cosh 2\alpha\}. \quad . \quad . \quad . \quad (15.27)$$

Next consider

$$\Omega_1(z)=\mathrm{G}(\zeta z-c\sinh\zeta),\quad \omega_1(z)=0, \quad . \quad . \quad . \quad (15.28)$$

where G is a real constant, whence

$$\Omega_1{}'(z)=\mathrm{G}\zeta,\quad \Omega_1{}''(z)=\frac{\mathrm{G}}{c\sinh\zeta};\quad \Omega_1{}'''(z)=-\frac{\mathrm{G}\cosh\zeta}{c^2\sinh^2\zeta}.$$

From (9.12) we find

$$2\mu w_0=\tfrac{1}{2}h^2\mathrm{G}(\zeta+\bar{\zeta})-\tfrac{1}{8}(1-\eta)\mathrm{G}\{z\bar{z}(\zeta+\bar{\zeta})-c^2\sinh(\zeta+\bar{\zeta})\}$$

and so, using (15.17), we have a boundary displacement given by

$$2\mu w_0=\mathrm{G}h^2\alpha-\tfrac{1}{8}(1-\eta)\mathrm{G}\left\{2\alpha\left(c^2\cosh 2\alpha-\frac{a^2b^2}{\varpi^2}\right)-c^2\sinh 2\alpha\right\}. \quad (15.29)$$

From (10.4), (10.5) and (15.28) the stress couples are given by

$$\Lambda=\tfrac{1}{3}h^2(1+\eta)\mathrm{G}\zeta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15.30)$$

$$\Gamma=\tfrac{1}{6}h^2(1-\eta)\mathrm{G}\frac{\cosh\zeta}{\sinh\bar{\zeta}}+\frac{1}{15}(2-\eta)\mathrm{G}\frac{h^4}{c^2}\frac{\cosh\bar{\zeta}}{\sinh^3\bar{\zeta}}. \quad . \quad . \quad (15.31)$$

leading to a boundary flexural couple $\widetilde{\xi\eta}$ given by

$$4\widetilde{\xi\eta}=\tfrac{2}{3}h^2(1+\eta)\mathrm{G}\alpha+\tfrac{1}{3}h^2(1-\eta)\mathrm{G}\frac{\varpi^2}{ab}+\frac{2}{15}(2-\eta)\mathrm{G}\frac{\varpi^4h^2}{a^3b^3}. \quad (15.32)$$

Clearly from (15.31) Γ becomes infinite at the foci, where $\zeta=0$ and $i\pi$. Now take

$$\Omega_1(z)=0,\quad \omega_1(z)=\tfrac{1}{2}\mathrm{G}c^4(\zeta-\tfrac{1}{2}\sinh 2\zeta). \quad . \quad . \quad . \quad (15.33)$$

Then, from (9.12),

$$2\mu w_0=\frac{1}{32}\mathrm{G}c^2\{\sinh 2\zeta+\sinh 2\bar{\zeta}-(2\zeta+2\bar{\zeta})\},$$

and so, from (15.17), the boundary displacement is given by

$$2\mu w_0 = \frac{1}{16}(1-\eta)\mathrm{G}\left\{\sinh 2\alpha\left(c^2\cosh 2\alpha - \frac{2a^2b^2}{\varpi^2}\right) - c^2 2\alpha\right\}. \qquad (15.34)$$

The stress couples are given by

$$\mathit{\Lambda}=0, \quad \mathit{\Gamma} = -\tfrac{1}{6}h^2(1-\eta)\mathrm{G}\coth\bar{\zeta}, \quad . \quad . \quad . \quad . \qquad (15.35)$$

and the boundary flexural couple by

$$4\widetilde{\xi\eta} = -\tfrac{1}{3}(1-\eta)\mathrm{G}\frac{\varpi^2h^2}{ab}. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (15.36)$$

Again, from (15.35), $\mathit{\Gamma}$ becomes infinite at the foci, but since there $\zeta = \pm\bar{\zeta}$, this is wiped out by the corresponding first term in (15.31), when the solutions corresponding to the complex potentials (15.28) and (15.33) are combined.

Finally, consider

$$\mathit{\Omega}_1(z)=0, \quad \omega_1(z) = \mathrm{H}c^2\zeta. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (15.37)$$

This gives a boundary displacement

$$2\mu w_0 = -\tfrac{1}{4}(1-\eta)\mathrm{H}c^2\alpha, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (15.38)$$

whilst the stress couples are given by

$$\mathit{\Lambda}=0, \quad \mathit{\Gamma} = -\tfrac{1}{6}(1-\eta)h^2\mathrm{H}\frac{\cosh\bar{\zeta}}{\sinh^3\bar{\zeta}} \quad . \quad . \quad . \quad . \qquad (15.39)$$

and becomes infinite at the foci, but these annul the corresponding second term in (15.31) if we choose H so that

$$\mathrm{H} = \frac{2}{5}\frac{2-\eta}{1-\eta}\mathrm{G}\frac{h^2}{c^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (15.40)$$

From (15.39) and (15.14) the boundary flexural couple is given by

$$4\widetilde{\xi\eta} = -\frac{2}{15}(2-\eta)\mathrm{G}\frac{h^4\varpi^4}{a^3b^3}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (15.41)$$

We are now in a position to calculate the flexural clamping couple $\widetilde{\xi\eta}$ round the boundary $\xi=\alpha$, when this edge is built-in or "horizontally" clamped. Combining the results (15.21), (15.24), (15.27) and using (15.8)–(15.11), we find, in terms of the actual deflexion $-\mathrm{W}$ given by (15.20) and the plate modulus P of (11.5)

$$\widetilde{\xi\eta} = -\frac{8\mathrm{PW}}{\varpi^2} + \frac{ph}{40}\frac{8+\eta+\eta^2}{1-\eta} + \frac{6}{5}(8+\eta)\mathrm{PW}h^2\frac{(a^2+b^2)}{a^4b^4}\{2\varpi^2-a^2-b^2\}, \quad . \quad . \quad . \qquad (15.42)$$

of which only the first term is that corresponding to the elementary theory. The general result appears to be new.

We go on to consider the elliptic plate under *constant flexural couple only* round the boundary, *i. e.* the case $p=0$ and $\widetilde{\xi\eta}=\text{const.}=\mathrm{M}$ along $\xi=\alpha$, the solution of which appears in none of the textbooks. Taking the complex potentials

$$\Omega_1(z)=\mathrm{G}(\zeta z-c\sinh\zeta)+\mathrm{B}z; \quad \ldots \ldots \quad (15.43)$$

$$\omega_1(z)=\tfrac{1}{4}\mathrm{G}c^2(2\zeta-\sinh 2\zeta)+\frac{2}{5}\,\frac{2-\eta}{1-\eta}\,\mathrm{G}h^2\zeta+\mathrm{F}, \quad \ldots \quad (15.44)$$

we find on combining the corresponding results for the boundary displacements

$$2\mu w_0=h^2(\mathrm{B}+\mathrm{G}\alpha)-\tfrac{1}{4}(1-\eta)(\mathrm{B}+\mathrm{G}\alpha-\tfrac{1}{2}\mathrm{G}\sinh 2\alpha)\left(c^2\cosh 2\alpha-\frac{a^2b^2}{\varpi^2}\right)$$
$$-\tfrac{1}{4}(1-\eta)\left\{\mathrm{F}+\frac{2}{5}\,\frac{2-\eta}{1-\eta}\,\mathrm{G}h^2\alpha+\tfrac{1}{8}\mathrm{G}c^2[4\alpha+\sinh 4\alpha-4\sinh 2\alpha]\right\}, \quad (15.45)$$

and from the results for the boundary flexural couples

$$4\mathrm{M}=\tfrac{2}{3}h^2(1+\eta)(\mathrm{B}+\mathrm{G}\alpha). \quad \ldots \ldots \quad (15.46)$$

If w_0 vanishes round the boundary, the vanishing of the term in ϖ^{-2} makes

$$\mathrm{B}+\mathrm{G}\alpha=\tfrac{1}{2}\mathrm{G}\sinh 2\alpha$$

and, using (15.46), we then find

$$\mathrm{G}=\frac{12\mathrm{M}\,\mathrm{cosech}\,2\alpha}{h^2(1+\eta)}, \quad \ldots \ldots \quad (15.47)$$

$$\mathrm{B}=\frac{6\mathrm{M}}{h^2(1-\eta)}(1-2\alpha\,\mathrm{cosech}\,2\alpha), \quad \ldots \ldots \quad (15.48)$$

$$\mathrm{F}=\frac{24\mathrm{M}}{1-\eta^2}\{1-\tfrac{1}{5}(2-\eta)\alpha\,\mathrm{cosech}\,2\alpha\}-\frac{3\mathrm{M}c^2}{(1+\eta)h^2}\{\cosh 2\alpha-2+2\alpha\,\mathrm{cosech}\,2\alpha\}. \quad \ldots \quad (15.49)$$

From (15.20) the central deflexion W is given by

$$\mathrm{W}=\frac{\mathrm{M}c^2}{4(1+\eta)\mathrm{P}}\{\cosh 2\alpha-2+2\alpha\,\mathrm{cosech}\,2\alpha\}-\frac{2}{5}\,\frac{\mathrm{M}h^2}{\mathrm{P}}\,\frac{8+\eta}{1-\eta^2}\,\alpha\,\mathrm{cosech}\,2\alpha. \quad \ldots \quad (15.50)$$

Since, as $\alpha\rightarrow\infty$, $c\rightarrow 0$, $b\rightarrow a$, $c^2\cosh 2\alpha\rightarrow 2a^2$, and $\alpha\,\mathrm{cosech}\,2\alpha\rightarrow 0$, we find that (15.43), (15.44), (15.47)–(15.49) lead satisfactorily to the results (14.15) for the circular plate, whilst (15.50) leads to (14.14).

Finally, in order to obtain the solution for the elliptic plate under *pressure* p *and a constant flexural couple* M round the boundary, we combine the solutions given by all the complex potentials considered in this section, *i. e.*, we have a boundary displacement w_0 given by combining (15.19), (15.23), (15.26), (15.29), (15.34), (15.38), and a boundary flexural couple $\widetilde{\xi\eta}$ given by combining (15.21), (15.24), (15.27), (15.32), (15.36), and (15.41). The coefficient of ϖ^{-4} in $2\mu w_0$ vanishes if

$$\mathrm{A}\cosh 2\alpha+\mathrm{C}\cosh 4\alpha+\frac{3p}{32h^3}=0,$$

and the coefficient of ϖ^{-2} in $4\widetilde{\xi\eta}$ vanishes if

$$4(1-\eta)\left\{\mathrm{A}\cosh 2\alpha+\mathrm{C}\cosh 4\alpha+\frac{3p}{32h^3}\right\}-4\left\{\mathrm{A}\cosh 2\alpha+\frac{1}{8}\,\frac{p}{h^3}\right\}=0,$$

whence

$$\mathrm{A}=-\frac{p}{8h^3}\,\mathrm{sech}\,2\alpha, \quad \mathrm{C}=\frac{p}{32h^3}\,\mathrm{sech}\,4\alpha. \quad \ldots \quad (15.51)$$

Using these values of A and C, the coefficient of ϖ^2 in $4\widetilde{\xi\eta}$ vanishes if

$$\mathrm{E}=\frac{3pc^2}{16h^3}\frac{\sinh^2 2\alpha}{\cosh 4\alpha}-\frac{3p}{20h}\,\frac{2-\eta}{1-\eta}\operatorname{sech} 2\alpha. \quad . \quad . \quad . \quad (15.52)$$

Next, the vanishing of the coefficient of ϖ^{-2} in $2\mu w_0$ leads to

$$\mathrm{G}\sinh 2\alpha-2(\mathrm{B}+\mathrm{G}\alpha)=\frac{3p}{10h}\,\frac{\eta^2+9\eta-2}{1-\eta^2}+\frac{pc^2}{8h^3}\left\{\frac{\cosh 4\alpha}{\cosh 2\alpha}-\frac{\cosh 2\alpha}{\cosh 4\alpha}\right\}, \quad (15.53)$$

whilst the remaining term in $4\widetilde{\xi\eta}$ gives, if we set $\widetilde{\xi\eta}=\text{const.}=\mathrm{M}$,

$$\begin{aligned}\mathrm{B}'=\mathrm{B}+\mathrm{G}\alpha=\frac{3p}{20h}\,\frac{2+\eta}{1+\eta}+\frac{6\mathrm{M}}{(1+\eta)h^2}-\frac{3}{16}\,\frac{1-\eta}{1+\eta}\,\frac{pc^2}{h^3}\,\frac{\sinh^4 2\alpha}{\cosh 2\alpha\cosh 4\alpha}&\\ -\frac{3}{16}\,\frac{pc^2}{h^3}\,\frac{\sinh^2 2\alpha}{\cosh 2\alpha}.& \quad . \quad . \quad . \quad (15.54)\end{aligned}$$

From (15.53) and (15.54) we find

$$\mathrm{G}\sinh 2\alpha=\frac{12p}{5h}\,\frac{\eta}{1-\eta^2}+\frac{12\mathrm{M}}{(1+\eta)h^2}-\frac{1}{8}\,\frac{5-\eta}{1+\eta}\,\frac{pc^2}{h^3}\,\frac{\sinh^4 2\alpha}{\cosh 2\alpha\cosh 4\alpha}, \quad (15.55)$$

and we now have the necessary constants of the complex potentials which affect the stresses, since we can rewrite (15.43) as

$$\Omega_1(z)=\mathrm{G}(\zeta-\alpha)z-\mathrm{G}c\sinh\zeta+\mathrm{B}'z$$

and use G from (15.55) and B′ from (15.54). The vanishing of the remaining term in $2\mu w_0$ gives

$$\left.\begin{aligned}\mathrm{F}=\frac{6\mathrm{M}}{h^2(1+\eta)}&\left\{\frac{4h^2}{1-\eta}+c^2\left(1-\frac{1}{2}\cosh 2\alpha-\alpha\operatorname{cosech} 2\alpha\right)-\frac{4}{5}h^2\frac{2-\eta}{1-\eta}\alpha\operatorname{cosech} 2\alpha\right\}\\ &+\frac{3ph}{5(1-\eta^2)}\left\{2+\eta-\frac{8}{5}\eta\frac{2-\eta}{1-\eta}\alpha\operatorname{cosech} 2\alpha\right\}\\ &+\frac{3pc^2}{40h(1-\eta^2)}\left\{(\eta^2-4\eta-7)\cosh 2\alpha-(\eta^2+9\eta-12)\operatorname{sech} 2\alpha\right.\\ &\qquad\left.-5(1-\eta)\frac{\cosh 2\alpha}{\cosh 4\alpha}-\frac{2}{3}(2-\eta)(5-\eta)\frac{\alpha\sinh^3 2\alpha}{\cosh 2\alpha\cosh 4\alpha}\right\}\\ &+\frac{3pc^2}{64h^3}\,\frac{5+\eta}{1-\eta}\,\frac{\sinh^4 2\alpha}{\cosh 4\alpha}-\frac{pc^2}{16h^3}\,\frac{5-\eta}{1+\eta}\,\frac{\sinh^4 2\alpha}{\cosh 2\alpha\cosh 4\alpha}.\end{aligned}\right\} \quad . \quad . \quad . \quad (15.56)$$

For the central deflexion, (15.20) gives, on using (15.54)–(15.56),

$$\left.\begin{aligned}2\mu\mathrm{W}=3\left(\frac{1-\eta}{1+\eta}\right)\mathrm{M}&\left\{\frac{1}{4}\frac{c^2}{h^2}(\cosh 2\alpha-2+2\alpha\operatorname{cosech} 2\alpha)-\frac{2}{5}\,\frac{8+\eta}{1-\eta}\alpha\operatorname{cosech} 2\alpha\right\}\\ &-\frac{6}{25}ph\eta\frac{(8+\eta)}{1-\eta^2}\alpha\operatorname{cosech} 2\alpha\\ &-\frac{3}{160}\,\frac{pc^2}{h}\left\{\frac{\eta^2+\eta+8}{1+\eta}\cosh 2\alpha-\frac{\eta^2+9\eta+8}{1+\eta}\operatorname{sech} 2\alpha\right.\\ &\qquad\left.+\frac{2}{3}\,\frac{(5-\eta)(8+\eta)}{1+\eta}\,\frac{\alpha\sinh^3 2\alpha}{\cosh 2\alpha\cosh 4\alpha}\right\}\\ &-\frac{3}{256}\,\frac{pc^4}{h^3}\,\frac{(5+\eta)(1-\eta)}{1+\eta}\,\frac{\sinh^4 2\alpha}{\cosh 4\alpha}\\ &+\frac{1}{64}\,\frac{pc^4}{h^3}\,\frac{(5-\eta)(1-\eta)}{1+\eta}\,\frac{\sinh^4 2\alpha}{\cosh 2\alpha\cosh 4\alpha}.\end{aligned}\right\} \quad . \quad . \quad . \quad (15.57)$$

As $\alpha\rightarrow\infty$, A, E, C, G$\rightarrow 0$, whilst B$'$ will be found to be given by (14.8) for the circular plate as it should. Again (15.56) and (15.57) become, as $\alpha\rightarrow\infty$, (14.9) and (14.13)–(14.14) respectively, in complete agreement with the results for the circular plate. Setting $p=0$ in (15.56), (15.57) also gives agreement with the results (15.49), (15.50) of the simpler problem above.

The result for the *simply supported* elliptic plate is given by putting M$=0$ in the above, and extends Galerkin's result for the corresponding problem of the elementary theory, of which Timoshenko (' Theory of Plates and Shells,' p. 292) quotes some numerical results.

16. *Extension to Variable Pressure and Gravitational Body Force.*

For brevity, the discussion of the semi-inverse method was here restricted to the case of zero body force and to constant pressure p. Gravitational body force in the direction of the negative Z-axis can be allowed for by replacing (7.7) and $\widehat{zz}$ by

$$(1+\eta)\chi=\Theta+\widehat{zz}-2\rho g(1+\eta)\mathrm{Z}, \qquad (16.1)$$

$$\widehat{zz}=\frac{p}{4h^3}\{\mathrm{Z}^3-3h^2\mathrm{Z}-2h^3\}+\frac{\rho g}{2h^2}\mathrm{Z}(\mathrm{Z}^2-h^2). \qquad (16.2)$$

The additions to our previous solutions (8.3), (8.8), (8.9), (9.3) and (9.8) are then given by

$$\Theta=-\rho gh+2\rho g(1+\eta)\mathrm{Z},\ \Phi=0,\ \Psi=0, \qquad (16.3)$$

$$\mathrm{D}=-2\rho ghz-\rho g(1+\eta)z\mathrm{Z}, \qquad (16.4)$$

$$2\mu w=\tfrac{1}{8}\rho g(1+\eta)z\bar{z}+\rho gh\mathrm{Z}+\tfrac{1}{2}\rho g(1-2\eta)\mathrm{Z}^2. \qquad (16.5)$$

As regards the case of variable pressure, it will be found on proceeding as before that the assumed form of $\widehat{zz}$ in (8.1), with p a function of z and $\bar{z}$, is only compatible with $\nabla_1{}^2 p=0$, so that p must necessarily be of the form

$$p=\tfrac{1}{2}\{\mathrm{P}''(z)+\bar{\mathrm{P}}''(\bar{z})\}. \qquad (16.6)$$

This is the only restriction on p if we widen the assumed form of Ψ in (8.1) to

$$\Psi=(\mathrm{Z}^2-h^2)(\psi_0+\psi_2\mathrm{Z}^2), \qquad (16.7)$$

leading to

$$\psi_0=-\tfrac{1}{2}\bar{\Omega}_1{}''(\bar{z})-\frac{3}{16h^3}\{\mathrm{P}'(z)+z\bar{\mathrm{P}}''(\bar{z})\}-\frac{1}{4h}\bar{\mathrm{P}}'''(\bar{z}), \qquad (16.8)$$

$$\psi_2=\frac{1}{8h^3}\bar{\mathrm{P}}'''(\bar{z}). \qquad (16.9)$$

The additional terms introduced by (16.6)–(16.9) are given by

$$\Theta=\frac{1}{4}\{\mathrm{P}''(z)+\bar{\mathrm{P}}''(\bar{z})\}+\left\{\frac{3}{16h^3}[z\bar{\mathrm{P}}'(\bar{z})+\bar{z}\mathrm{P}'(z)]+\frac{3}{8}[\mathrm{P}''(z)+\bar{\mathrm{P}}''(\bar{z})]\right\}\mathrm{Z}$$
$$-(2+\eta)\frac{1}{8h^3}\{\mathrm{P}''(z)+\bar{\mathrm{P}}''(\bar{z})\}\mathrm{Z}^2, \qquad (16.10)$$

$$\Phi=\frac{1}{2}z\bar{\mathrm{P}}'''(\bar{z})+\left\{(1-\eta)\frac{3}{16h^3}[\mathrm{P}(z)+\frac{1}{2}z^2\bar{\mathrm{P}}''(\bar{z})]+\frac{3}{8h}z\bar{\mathrm{P}}'''(\bar{z})\right\}\mathrm{Z}+\frac{1}{2}\bar{\mathrm{P}}''''(\bar{z})\mathrm{Z}^2$$
$$-\left\{\frac{1}{4h}\bar{\mathrm{P}}''''(\bar{z})+(2-\eta)\frac{1}{8h^3}z\bar{\mathrm{P}}'''(\bar{z})\right\}\mathrm{Z}^3+(3-\eta)\frac{1}{40h^3}\bar{\mathrm{P}}''''(\bar{z})\mathrm{Z}^5, \qquad (16.11)$$

$$\begin{aligned}8\mu D=&\frac{3}{2}P'(z)-\frac{1}{2}z\overline{P}''(\bar{z})+\left\{(1-\eta)\frac{3}{8h^3}[\bar{z}P(z)+\frac{1}{2}z^2\overline{P}'(\bar{z})]+\frac{3}{4h}[P'(z)+z\overline{P}''(\bar{z})]\right\}Z\\&+\overline{P}'''(\bar{z})Z^2-\left\{\frac{1}{4h^3}(2-\eta)[P'(z)+z\overline{P}''(\bar{z})]+\frac{1}{2h}\overline{P}'''(\bar{z})\right\}Z^3\\&-(3-\eta)\frac{1}{20h^3}\overline{P}'''(\bar{z})Z^5, \quad \ldots \ldots \ldots \quad (16.12)\end{aligned}$$

$$\begin{aligned}2\mu w=&\frac{3}{32h}[\bar{z}P'(z)+z\overline{P}'(\bar{z})]+\frac{1}{4}h[P''(z)+\overline{P}''(\bar{z})]-\frac{3}{128h^3}(1-\eta)[z^2\overline{P}(\bar{z})+\bar{z}^2P(z)]\\&-\frac{1}{4}[P''(z)+\overline{P}''(\bar{z})]Z-\left\{\frac{3}{32h^3}\eta[z\overline{P}'(\bar{z})+\bar{z}P'(z)]+\frac{3}{16h}[P''(z)+\overline{P}''(\bar{z})]\right\}Z^2\\&+\frac{1}{32h^3}(1+\eta)[P''(z)+\overline{P}''(\bar{z})]Z^4. \quad \ldots \ldots \ldots \quad (16.13)\end{aligned}$$

If we take $P(z)=\alpha z^3+\beta z^2$, for example, we have the case of uniformly varying pressure.

Conclusion.

In this paper the writer has sought to re-direct attention to the definition and nomenclature of stress resultants and stress couples in a plate; to show the usefulness of complex combinations of these stress resultants and couples in their transformation to other systems of co-ordinates, especially two dimensional orthogonal curvilinears, and to give the solution of the semi-inverse problem of plate theory in terms of complex potentials. The method is then illustrated by examples relative to various boundaries, some of the results being new.

References.

(1) Jeffreys, Phil. Mag. xxxii. p. 365 (1941).
(2) Stevenson, Proc. Roy. Soc. (Not yet published.)
(3) Kolossof, *Comptes Rendus*, clxxxiv. p. 512 (1927).
(4) Filon, Proc. Lond. Math. Soc. xliii. p. 152 (1937).

LXX. *The Temperature Distribution around a Spherical Hole in an Infinite Conducting Medium.*

By H. Ll. D. Pugh, B.Sc. and A. J. Harris, B.Sc.*

[Received April 16, 1942.]

Summary.

An integral expression is obtained for the temperature distribution in a conducting medium around a spherical hole given an arbitrary initial temperature distribution possessing spherical symmetry. The boundary

---

* Communicated by the Authors.

conditions assumed are that heat flows from the medium through the surface of the hole at a rate proportional to the surface temperature, and that there is no flow of heat at infinity.

The problem here considered is to find the temperature distribution after the lapse of time, t, in an infinite conducting medium containing a spherical hole when the medium loses heat through the surface of the hole at a rate proportional to the temperature of that surface. The initial temperature distribution is given and has spherical symmetry about the centre of the hole. It is assumed that there is no loss of heat through the boundary at infinity.

Let θ=temperature of a point in the medium.
r=distance of the point from the centre of the spherical hole.
a=radius of the spherical hole.
k=conductivity of the medium.
ρ=density of the medium.
S=specific heat of the medium.
K=emissivity of the surface of the hole.
t=time.
$h^2=\dfrac{k}{\rho\mathrm{S}}$=diffusivity of the medium.

The equation of heat conduction in three dimensions with spherical symmetry about the origin is

$$\frac{\partial}{\partial r}\left(kr^2\frac{\partial\theta}{\partial r}\right)-\rho\mathrm{S}r^2\frac{\partial\theta}{\partial t}=0,$$

which, when $r.\theta$ is taken as the dependent variable, may be written

$$\frac{\partial^2}{\partial r^2}(r.\theta)=\frac{1}{h^2}\frac{\partial}{\partial t}(r.\theta)\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

To solve the problem stated above, we must first obtain the solution of certain other cases with simpler initial conditions.

(*a*) Let the initial temperature distribution at $t=0$ be given by

$$\left.\begin{array}{l}\theta=0 \text{ for } a<r<r_1,\\ \theta=\theta_0 \text{ for } r_1<r\end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In accordance with Heaviside's operational calculus, where p^{-1} denotes the operation $\int_0^t dt$, we have on integrating (1)

$$r_1<r \qquad p^{-1}\frac{\partial^2(r.\theta)}{\partial r^2}=\frac{1}{h^2}(r.\theta-r.\theta_0)$$

or

$$\frac{\partial^2(r.\theta)}{\partial r^2}-\frac{p}{h^2}(r.\theta)+\frac{p}{h^2}(r.\theta_0)=0.\quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Writing q^2 for $\frac{p}{h^2}$ and solving the equation by treating q as a number we get

$$r.\theta=A_2e^{qr}+B_2e^{-qr}+r.\theta_0. \qquad (4)$$

For the range $a<r<r_1$, we have the same result except that θ_0 must be put equal to 0. Thus

$$r.\theta=A_1e^{qr}+B_1e^{-qr}. \qquad (5)$$

Boundary conditions

$$\left.\begin{array}{l} \text{At } r=a, \quad k\dfrac{\partial\theta}{\partial r}=K\theta \text{ for all } t. \\[2ex] \text{At } r=r_1, \ \theta \text{ and } \dfrac{\partial\theta}{\partial r} \text{ are continuous for all } t. \\[2ex] \text{At } r=\infty\ , \ r^2\dfrac{\partial\theta}{\partial r} \text{ must tend to zero.} \end{array}\right\} \qquad (6)$$

From (4) and (5)

$$k\left(A_1\left[\frac{qe^{qa}}{a}-\frac{e^{qa}}{a^2}\right]+B_1\left[-\frac{qe^{-qa}}{a}-\frac{e^{-qa}}{a^2}\right]\right)=K\left(A_1\frac{e^{qa}}{a}+B_1\frac{e^{-qa}}{a}\right). \qquad (7)$$

$$A_1\frac{e^{qr_1}}{r_1}+\frac{B_1e^{-qr_1}}{r_1}=\frac{A_2e^{qr_1}}{r_1}+\frac{B_2e^{-qr_1}}{r_1}+\theta_0. \qquad (8)$$

$$A_1\left[\frac{qe^{qr_1}}{r_1}-\frac{e^{qr_1}}{r_1^2}\right]+B_1\left[\frac{-qe^{-qr_1}}{r_1}-\frac{e^{-qr_1}}{r_1^2}\right]=A_2\left[\frac{qe^{qr_1}}{r_1}-\frac{e^{qr_1}}{r_1^2}\right]$$
$$+B_2\left[\frac{-qe^{-qr_1}}{r_1}-\frac{e^{-qr_1}}{r_1^2}\right]. \qquad (9)$$

$$\underset{r\to\infty}{\text{Limit}} \quad r^2\left\{A_2\left[\frac{qe^{qr}}{r}-\frac{e^{qr}}{r^2}\right]+B_2\left[\frac{-qe^{-qr}}{r}-\frac{e^{-qr}}{r^2}\right]\right\}=0. \qquad (10)$$

(10) cannot be satisfied unless $A_2=0$; the remaining equations then give

$$A_1=\frac{\theta_0(r_1q+1)}{2q}e^{-qr_1},$$

$$B_1=\frac{\theta_0(r_1q+1)}{2q}\frac{\left(qa-\frac{Ka}{k}-1\right)}{\left(qa+\frac{Ka}{k}+1\right)}e^{(2a-r_1)q}.$$

$$B_2=\frac{\theta_0e^{qr_1}}{2q}\left[1-r_1q+(qr_1+1)\frac{\left(qa-\frac{Ka}{k}-1\right)}{\left(qa+\frac{Ka}{k}+1\right)}e^{-2(r_1-a)q}\right].$$

Finally

$$a<r<r_1 \quad \theta=\frac{\theta_0(r_1q+1)e^{q(r-r_1)}}{2qr}$$
$$+\frac{\theta_0(r_1q+1)}{2qr}\frac{\left(qa-\frac{Ka}{k}-1\right)}{\left(qa+\frac{Ka}{k}+1\right)}e^{-(r+r_1-2a)q}. \qquad (11)$$

$$r_1<r \quad \theta=\theta_0+\frac{\theta_0(-r_1q+1)}{2qr}e^{-q(r-r_1)}$$

$$+\frac{\theta_0(r_1q+1)}{2qr}\frac{\left(qa-\frac{\mathrm{K}a}{k}-1\right)}{\left(qa+\frac{\mathrm{K}a}{k}+1\right)}e^{(r_1+r-2a)q}. \quad (12)$$

(*b*) Let the initial temperature distribution be

$$a<r<r_1 \quad \theta=0,$$
$$r_1<r<r_1+\delta r_1 \quad \theta=\theta_0,$$
$$r_1+\delta r_1<r \quad \theta=0.$$

The solution to this problem may be obtained by subtracting from the solution just found the corresponding solution when r_1 is put equal to $r_1+\delta r_1$:

$$a<r<r_1 \quad \theta=\frac{\theta_0}{2qr}[(r_1q+1)e^{q(r-r_1)}-([r_1+\delta r_1]q+1)e^{q(r-r_1-\delta r_1)}]$$

$$+\frac{\theta_0}{2qr}\left(\frac{qa-\frac{\mathrm{K}a}{k}-1}{qa+\frac{\mathrm{K}a}{k}+1}\right)[(r_1q+1)e^{-(r+r_1-2a)q}$$

$$-([r_1+\delta r_1]q+1)e^{-(r_1+\delta r_1+r-2a)q}],$$

$$r_1+\delta r_1<r \quad \theta=\frac{\theta_0}{2qr}[(-r_1q+1)e^{-q(r-r_1)}-(-[r_1+\delta r_1]q+1)e^{-q(r-r_1-\delta r_1)}]$$

$$+\frac{\theta_0}{2qr}\left(\frac{qa-\frac{\mathrm{K}a}{k}-1}{qa+\frac{\mathrm{K}a}{k}+1}\right)[(r_1q+1)e^{-(r_1+r-2a)q}$$

$$-([r_1+\delta r_1]q+1)e^{-(r_1+\delta r_1+r-2a)q}].$$

If we now proceed to the limit, making δr_1 tend to zero and θ_0 to infinity, so that the product $\theta_0\,\delta r_1$ tends to the limit I, we have

$$a<r<r_1 \quad \theta=\frac{\mathrm{I}r_1}{2r}\left[qe^{q(r-r_1)}+\frac{\left(qa-\frac{\mathrm{K}a}{k}-1\right)}{\left(qa+\frac{\mathrm{K}a}{k}+1\right)}qe^{\ q(r+r_1-2a)}\right]. \quad (13)$$

$$r_1<r \quad \theta=\frac{\mathrm{I}r_1}{2r}\left[qe^{-q(r-r_1)}+\frac{\left(qa-\frac{\mathrm{K}a}{k}-1\right)}{\left(qa+\frac{\mathrm{K}a}{k}+1\right)}qe^{-q(r+r_1\ \ 2a)}\right]. \quad (14)$$

(*c*) General distribution

$$a<r<\infty \quad \theta=f(r) \text{ at } t=0. \quad (15)$$

The general distribution may be solved by regarding it as the sum of a number of elements of the type discussed in (*b*). The solution is then the sum of the solutions for the separate elements or in the limit an integral of the solutions (13) and (14).

The element $\theta=f(r_1)$ in the range $r_1<r<r_1+\delta r_1$ gives a temperature distribution expressed by (13) and (14) in the ranges $0<r<r_1$ and $r_1<r$ respectively. The value of θ due to all such elements is at the point r

$$\theta=\int_a^r f(\lambda)\frac{\lambda}{2r}\left[qe^{-q(r-\lambda)}+\frac{\left(qa-\frac{\mathrm{K}a}{k}-1\right)}{\left(qa+\frac{\mathrm{K}a}{k}+1\right)}qe^{-q(r+\lambda-2a)}\right]d\lambda$$

$$+\int_r^\infty f(\lambda)\frac{\lambda}{2r}\left[qe^{q(r-\lambda)}+\frac{\left(qa-\frac{\mathrm{K}a}{k}-1\right)}{\left(qa+\frac{\mathrm{K}a}{k}+1\right)}qe^{-q(r+\lambda-2a)}\right]d\lambda. \quad (16)$$

Writing (16) in a form suitable for interpretation we get

$$\theta=\int_a^r\frac{f(\lambda)\lambda}{2r}qe^{-q(r-\lambda)}d\lambda+\int_r^\infty\frac{f(\lambda)\lambda}{2r}qe^{-q(\lambda-r)}d\lambda$$

$$+\int_a^\infty f(\lambda)\frac{\lambda}{2r}\left[q-2\left(\frac{\mathrm{K}a+k}{ka}\right)+\frac{2\left(\frac{\mathrm{K}a+k}{ka}\right)^2}{q+\left(\frac{\mathrm{K}a+k}{ka}\right)}\right]e^{-q(r+\lambda-2a)}d\lambda. \quad (17)$$

The interpretations of the operational expressions are as follows (see Jeffreys, H., ' Operational Methods in Mathematical Physics ').

$$qe^{-qx}=\frac{1}{h(\pi t)^{\frac{1}{2}}}e^{-\frac{x^2}{4h^2t}},$$

$$e^{-qx}=1-\mathrm{erf}\left(\frac{x}{2ht^{\frac{1}{2}}}\right),$$

$$\frac{1}{q+\alpha}e^{-qx}=\frac{1}{\alpha}\left\{1-\mathrm{erf}\frac{x}{2ht^{\frac{1}{2}}}-\exp\,(\alpha^2h^2t+\alpha x)\left[1-\mathrm{erf}\left(\frac{x}{2ht^{\frac{1}{2}}}+\alpha ht^{\frac{1}{2}}\right)\right]\right\},$$

where erf x is the error function defined as

$$\mathrm{erf}\,x=\frac{2}{\sqrt{\pi}}\int_0^x e^{-\lambda^2}d\lambda.$$

Then, inserting the interpretations in (17), we get

$$\theta=\frac{1}{2r}\int_a^r\frac{f(\lambda)\lambda}{h\pi^{\frac{1}{2}}t^{\frac{1}{2}}}e^{-\frac{(r-\lambda)^2}{4h^2t}}d\lambda+\frac{1}{2r}\int_r^\infty\frac{f(\lambda).\lambda}{h\pi^{\frac{1}{2}}t^{\frac{1}{2}}}e^{-\frac{(\lambda-r)^2}{4h^2t}}d\lambda$$

$$+\frac{1}{2r}\int_a^\infty f(\lambda)\lambda\left[\frac{1}{h\pi^{\frac{1}{2}}t^{\frac{1}{2}}}e^{-\frac{(r+\lambda-2a)^2}{4h^2t}}-2\left(\frac{\mathrm{K}a+k}{ka}\right)\left(1-\mathrm{erf}\frac{r+\lambda-2a}{2ht^{\frac{1}{2}}}\right)\right.$$

$$+2\left(\frac{\mathrm{K}a+k}{ka}\right)\left(1-\mathrm{erf}\frac{r+\lambda-2a}{2ht^{\frac{1}{2}}}-\exp\left\{\left(\frac{\mathrm{K}a+k}{ka}\right)(h^2t+r+\lambda-2a)\right\}\right.$$

$$\left.\left.\times\left\{1-\mathrm{erf}\left[\left(\frac{r+\lambda-2a}{2ht^{\frac{1}{2}}}\right)+\left(\frac{\mathrm{K}a+k}{ka}\right)ht^{\frac{1}{2}}\right]\right\}\right)\right]d\lambda. \quad (18)$$

The first two integrals form a single integral from a to ∞. When the substitution $\dfrac{r-\lambda}{2h\sqrt{t}}=\beta$ is made in the first two integrals and $\dfrac{r+\lambda-2a}{2h\sqrt{t}}=\beta$ in the last integral, (18) becomes

$$\theta=\frac{1}{r\sqrt{\pi}}\int_{-\frac{r-a}{2h\sqrt{t}}}^{\infty} f(r+2h\sqrt{t}\beta).(r+2h\sqrt{t}\beta)e^{-\beta^2}d\beta$$

$$+\frac{1}{r\sqrt{\pi}}\int_{+\frac{r-a}{2h\sqrt{t}}}^{\infty} f(2a-r+2h\sqrt{t}\beta).(2a-r+2h\sqrt{t}\beta)e^{-\beta^2}d\beta$$

$$-2\left(\frac{\mathrm{K}a+k}{ka}\right)\frac{h\sqrt{t}}{r}e^{\left(\frac{\mathrm{K}a+k}{ka}\right)^2h^2t}\int_{+\frac{r-a}{2h\sqrt{t}}}^{\infty} f(2a-r+2h\sqrt{t}\beta)$$

$$\times(2a-r+2h\sqrt{t}.\beta)\left[1-\mathrm{erf}\left(\beta+\frac{\mathrm{K}a+k}{ka}h\sqrt{t}\right)\right]e^{\left(\frac{\mathrm{K}a+k}{ka}\right)2h\sqrt{t}.\beta}.d\beta.$$

Road Research Laboratory.
March 1942.

LXXI. *The Damping of the Heaving and Pitching Motion of a Ship.*

By T. H. Havelock, F.R.S.*

[Received June 12, 1942.]

1. In a recent paper (1940) I discussed the damping of the rolling of a ship in still water due to the radiation of energy in the wave motion set up by the rolling. The following note is a similar examination of heaving and pitching oscillations; an attempt is made to estimate the dissipation of energy in wave motion and comparison is made with such experimental results as are available.

The problem may be stated first in relation to heaving motion. Consider a body of mass M floating freely in water, and suppose it is acted on by a periodic force E cos pt and is making small vertical oscillations; let ζ be the vertical displacement upwards from the equilibrium position. The equation of motion, for a frictionless liquid, is

$$\mathrm{M}\ddot{\zeta}=\mathrm{X}-\mathrm{M}g+\mathrm{E}\cos pt, \qquad \ldots\ldots\ (1)$$

where X is the vertical resultant of the fluid pressures on the immersed surface. For an exact solution we should have to determine the velocity potential of the fluid motion so as to satisfy the boundary condition at the moving surface of the solid and the condition of constant pressure over the free surface of the liquid. Failing such a solution we proceed by approximations. One part of the force X-Mg is the additional hydrostatic buoyancy, $g\rho S\zeta$ upwards, assuming the solid to have vertical sides near the water-line and S to be the area of the water plane section. Suppose now that the motion of the body is a forced vibration of frequency p and that the energy radiated is relatively small; then, as in similar problems, for instance the scattering of sound waves by a movable obstacle, it is assumed that the rest of the resultant fluid pressure may be expressed as the sum of two terms, one proportional to $\ddot{\zeta}$ and the other to $\dot{\zeta}$. The factor of the first term represents the so-called added mass, while the second corresponds to the loss of energy by propagation of waves outwards. In these circumstances, equation (1) is reduced to the form

---

* Communicated by the Author.

$$M'\ddot{\zeta}+N\dot{\zeta}+g\rho S\zeta=E\cos pt, \quad . \; . \; . \; . \; . \; . \; . \quad (2)$$

where M′ is the total effective mass.

2. Various empirical formulæ have been devised for the effective mass of a ship for heaving motion, and for flexural vibrations. Reference may be made in particular to Lewis (1929) for ship forms, and to Browne, Moullin and Perkins (1930) for the added mass of prisms floating in water. The basic assumption in these studies is to neglect the wave disturbance and to suppose the fluid motion to be that due to a certain solid moving in an infinite liquid, the solid being made up of the immersed part of the floating body and its reflexion in the free surface of the water. The experiments of Browne, Moullin and Perkins showed that this leads to a reasonable value of the added mass, the calculated values being rather higher than those deduced from the experiments.

It is the second term of equation (2), namely $N\dot{\zeta}$, with which the present paper is specially concerned. Instead of calculating the fluid pressures, an alternative method is to work out the mean rate of propagation of energy outwards in the wave motion, and equating this to the mean value of $N\dot{\zeta}^2$ we obtain a value for N for the given frequency. This procedure is permissible under the assumed conditions under which the motion is a forced simple harmonic vibration and the radiated energy is small. To obtain the corresponding logarithmic decrement for the damped natural vibrations, these may be taken as approximately of period $2\pi/\sigma$, with

$$\sigma^2 M'=g\rho S. \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (3)$$

Then the logarithmic decrement is given by $\pi N/\sigma M'$, with N having its value for the frequency σ. There is very little work, theoretical or experimental, to which reference can be made. Browne, Moullin and Perkins (1930) measured the damping for prisms vibrating in air and when immersed in water; they conclude "The damping added by the water is negligible compared with the damping due to the supports, a result which might not have been expected." But in those experiments the prisms were not floating freely and the frequency was of the order of 13 per second; it can readily be estimated that the energy in the wave motion would then be very small. However, the experiments show that damping by fluid friction and eddies was also negligible. Reference may be made specially to work by Schuler (1936) with a vibrating prism of rectangular section, in which direct measurement was made of the amplitude of the waves. The logarithmic decrement was also measured, and it was concluded from the dimensional form of the results that the damping was due to wave motion, viscous and other damping being negligible in comparison. Schuler gives no theoretical calculation of the damping, and unfortunately the data necessary for making a comparison with theory are not recorded, such as the effective mass and restoring force and the free or forced nature of the vibrations.

Coming to the ship problem, as far as published work is concerned there is practically no accurate information about the damping of natural heaving. It is usually stated to be very large, any natural vibrations dying out very quickly. The only numerical estimate appears to be that given by Horn (1936) and said to be an average result derived from a large number of models; his estimate gives a logarithmic decrement for natural heaving of about 1·45. This is very large, and would mean that the amplitude is reduced by about one-half in each swing. It is also stated that the decrement for natural pitching oscillations is of the same order of magnitude.

3. We now examine the waves produced by an oscillating body, and we adopt the method of replacing it by some suitable distribution of alternating sources.

We consider first two-dimensional fluid motion, and we take the origin O in the free surface, Ox horizontal, and Oz vertically upwards. If there is a source of strength $m\cos pt$ per unit length at a depth f, that is at the point $(0, -f)$, the velocity potential is given by

$$\phi=-me^{ipt}\log\frac{r_1}{r_2}+2me^{ipt}\int_0^\infty\frac{e^{-\kappa(f-z)}\cos\kappa x}{\kappa-\kappa_0+i\mu}\,d\kappa, \quad . \quad . \quad . \quad . \quad (4)$$

where $p^2=g\kappa_0$, $r_1{}^2=x^2+(z+f)^2$, $r_2{}^2=x^2+(z-f)^2$; and we take the limiting value of the real part of the expression as μ is made zero.

This leads to a surface elevation given by

$$\zeta=\frac{2\pi mp}{g}e^{-\kappa_0 f}\cos(pt\mp\kappa_0 x)-\frac{2mp}{g}\sin\sigma t\int_0^\infty\frac{\kappa\cos\kappa f-\kappa_0\sin\kappa f}{\kappa^2+\kappa_0{}^2}e^{\mp\kappa x}\,dk, \quad (5)$$

where the upper or lower signs are taken according as x is positive or negative.

The first term in (5) gives the regular waves propagated outwards on either side; if A is the amplitude of these waves and E the mean rate of propagation of energy outwards per unit length, we have, taking account of both sides of the origin,

$$\mathrm{E}=g\rho p\mathrm{A}^2/2\kappa_0=2\pi^2m^2\rho pe^{-2\kappa_0 f}. \quad . \quad . \quad . \quad . \quad . \quad (6)$$

By summation, or integration, we can obtain the corresponding expression for any given distribution of periodic sources in the liquid.

4. Consider a long prism, of rectangular cross-section and of breadth $2b$, immersed in water to a depth f and made to perform small vertical oscillations $a \sin pt$. For an approximate solution we suppose the motion to be two-dimensional and to be that due to a uniform distribution of sources, of density $(pa/2\pi) \cos pt$, over the immersed base of the prism at its mean depth f. The regular waves on the side $x>0$ are given, from (5), by

$$\zeta=(p^2a/g)e^{-\kappa_0 f}\int_{-b}^{b} \cos \{pt-\kappa_0(x-h)\}\,dh$$

$$=2ae^{-\kappa_0 f} \sin (\kappa_0 b) \cos (pt-\kappa_0 x). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

Hence, for the mean rate of radiation of energy per unit length of the prism, we have

$$\mathrm{E}=(2g\rho pa^2/\kappa_0)e^{-2\kappa_0 f} \sin^2 (\kappa_0 b). \quad . \quad . \quad . \quad . \quad . \quad (8)$$

If the wave-length $2\pi/\kappa_0$ is large compared with the breadth of the prism, we have the simpler forms

$$\zeta=2\kappa_0 abe^{-\kappa_0 f}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

$$\mathrm{E}=2g\rho p\kappa_0 a^2b^2e^{-2\kappa_0 f}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

In the experiments by Schuler (1936) a rectangular prism was used and the amplitude of the waves and other quantities measured directly. Schuler obtained the expression (7) by an indirect energy method suggested by Prandtl, and it was contrasted with the source theory of the effect; however, we have seen that it follows from assuming a uniform distribution of sources over the base of the prism. The interesting point is that the experimental results agree reasonably well with the expression (7) for periods such that the wave-length $2\pi/\kappa_0$ is greater than the breadth $2b$ of the prism.

5. We now apply these results to the heaving of a ship in still water. We may, as in similar cases, treat the motion as two-dimensional in the first instance, an approximation which may be supported by the experiments of Browne, Moullin and Perkins. These authors also give an approximate formula for the added mass of a ship of normal form in vertical heaving motion; this is given as $0{\cdot}95\rho b^2 l$, where ρ is the density of water, $2b$ the maximum beam and l the total length of the ship.

We take an example from recent work by Kent and Cutland (1941), carried out on models at the National Physical Laboratory. The data for the corresponding ship are: length 400 ft., beam 55 ft., draught 24 ft., displacement 11,332 tons, natural heaving period 7·42 sec. From the formula just given the added mass comes out as 8200 tons; thus the total effective mass M′ in equation (2) is about 20,000 tons. It is of interest to check this result in a different way. If $2\pi/\sigma$ is the period when damping is neglected, we have the relation given in (3). The change in period due to damping is relatively small, so we may use the recorded period; further, estimating the water plane area S as 17,600 sq. ft., we obtain from (3) a value for M′ of about 20.000 tons.

Suppose now that the heaving motion is given by $\zeta = a \sin \sigma t$. The wave-length for a period of 7·42 sec. is about five times the beam of the ship. It is thus permissible to take a simple distribution, namely a uniform line source, of strength $m \cos \sigma t$ per unit length, extending over the length L of the ship at some suitable mean depth f. We take the value of m to correspond to the rate of alteration in displaced volume of the ship, which is $S\dot{\zeta}$ or $Sa\sigma \cos \sigma t$. Hence we take

$$m = S\sigma a / 2\pi L. \qquad \qquad (11)$$

We put this value of m in (6) and, using

$$\tfrac{1}{2} N \sigma^2 a^2 = EL, \qquad \qquad (12)$$

we obtain

$$N = (\rho\sigma S^2/L) e^{-2\kappa_0 f}. \qquad \qquad (13)$$

For the corresponding logarithmic decrement we have

$$\delta = \frac{\pi N}{\sigma M'} = \frac{\pi \rho S^2}{M'L} e^{-2\sigma^2 f/g}. \qquad \qquad (14)$$

Putting in the values already given for this case and taking $f = 20$ ft. as a mean depth, (14) gives the value $\delta = 1{\cdot}4$. The agreement with Horn's estimate is, of course, merely a coincidence. Clearly, this large value of δ goes beyond the assumption on which N has been calculated, namely that the damping is small enough to allow approximately simple harmonic waves to be established. Nevertheless the calculation is sufficient to show that wave motion is quite adequate to account for the large damping which has been observed in practice.

6. There does not seem to have been any experimental work on cases of three-dimensional fluid motion. We shall examine the corresponding theory, as it will allow of more detailed calculation for other source distributions and also of application to pitching. Consider a point source of alternating strength $m \cos pt$ at a depth f below the surface, that is, with the source at the point $(0, 0, -f)$. In this case the surface elevation is given by

$$\zeta = \frac{2ipm}{g} e^{ipt} \left\{ \frac{1}{(r^2+f^2)^{\frac{1}{2}}} - \frac{2\kappa_0}{\pi} \int_0^\infty \int_0^\infty \frac{\kappa \sin \kappa f - \kappa_0 \cos \kappa f}{\kappa^2 + \kappa_0^2} e^{-\kappa r \cosh u} \, du \, d\kappa \right.$$
$$\left. - i\pi\kappa_0 e^{-\kappa_0 f} H_0^{(2)}(\kappa_0 r) \right\}, \qquad \qquad (15)$$

where $r^2 = x^2 + y^2$, $p^2 = g\kappa_0$, $H_0^{(2)} = J_0 - iY_0$, and the real part of the expression is to be taken. There is a corresponding expression for the velocity potential. The first two terms in (15) represent local standing oscillations of the surface, and the third term the symmetrical circular waves propagated outwards. For the present purpose we only require the wave motion at a great distance, and the first term in the asymptotic expansion of the Bessel functions is sufficient; hence, for this part of the velocity potential and surface elevation, we obtain

$$\phi \sim 2\pi\kappa_0 m \left(\frac{2}{\pi \kappa_0 r} \right)^{\frac{1}{2}} e^{-\kappa_0 f + \kappa_0 z} \sin\left(pt + \frac{\pi}{4} - \kappa_0 r \right). \qquad \qquad (16)$$

$$\zeta \sim \frac{2\pi\kappa_0 pm}{g}\left(\frac{2}{\pi\kappa_0 r}\right)^{\frac{1}{2}} e^{-\kappa_0 f}\cos\left(pt+\frac{\pi}{4}-\kappa_0 r\right). \quad . \quad . \quad . \quad (17)$$

The rate of transmission of energy outwards is obtained from the rate of work of the fluid pressure over the surface of a vertical cylinder of radius r, that is, from

$$-\int_{-\infty}^{0}\rho\frac{\partial\phi}{\partial t}\frac{\partial\phi}{\partial r}2\pi r\,dz. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

Using (16), we obtain for E, the mean rate of transmission of energy outwards,

$$\mathrm{E}=4\pi^2\rho\kappa_0 pm^2 e^{-2\kappa_0 f}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

This result may be generalized to cover any given distribution over a surface S in the liquid. Let $m\cos pt$ be the source strength per unit area at a point (x', y', z') on this surface; we have to substitute for r in (15) the quantity

$$(r^2-2rx'\cos\theta-2ry'\sin\theta+x'^2+y'^2)^{\frac{1}{2}}, \quad . \quad . \quad . \quad (20)$$

and then integrate over the distribution.

It is readily seen that we only need the approximation for r large, namely

$$\phi\sim\left(\frac{8\pi\kappa_0}{r}\right)^{\frac{1}{2}}e^{\kappa_0 z}\left\{\mathrm{P}\sin\left(pt+\frac{\pi}{4}-\kappa_0 r\right)+\mathrm{Q}\cos\left(pt+\frac{\pi}{4}-\kappa_0 r\right)\right\}, \quad (21)$$

where

$$\mathrm{P}+i\mathrm{Q}=\iint m(x', y', z')e^{\kappa_0 z'+i\kappa_0(x'\cos\theta+y'\sin\theta)}\,d\mathrm{S}. \quad . \quad . \quad . \quad (22)$$

From these expressions, we obtain for the mean rate of outflow of energy

$$\mathrm{E}=2\pi\rho\kappa_0 p\int_0^{2\pi}(\mathrm{P}^2+\mathrm{Q}^2)\,d\theta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

7. Consider a circular cylinder, of radius b, immersed with its axis vertical to a depth f and making forced vertical oscillations given by $a\sin pt$. As in the two-dimensional problem, if $2\pi g/p^2$ is larger than the diameter of the cylinder, we may assume the wave motion to be due to an alternating source $m\cos pt$ at a depth f, with $4\pi m=\pi b^2 pa$. Hence, from (19), we have

$$\mathrm{E}=\frac{\pi^2\rho}{4g}b^4a^2p^5e^{-2p^2f/g}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

Assuming that this may be used to evaluate N for the natural damped vibrations when the cylinder is floating freely, we obtain

$$\mathrm{N}=\frac{\pi^2\rho}{2g}b^4\sigma^3e^{-2\sigma^2f/g}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

From the usual hydrostatic theory, $\sigma^2=g/f$. Without attempting to evaluate the effective mass in this case, we write $\mathrm{M}'=(1+\mu)\mathrm{M}=(1+\mu)\pi\rho b^2 f$. Hence, with these values, the logarithmic decrement is

$$\delta=\frac{\pi\mathrm{N}}{\sigma\mathrm{M}'}=\frac{\pi^2b^2e^{-2}}{(1+\mu)f^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

For instance, with $f=4b$ and neglecting μ we should have $\delta=0{\cdot}04$.

8. We may now attempt to apply these results to the pitching motion of a ship. For a long narrow ship the appropriate source distribution could be taken over the longitudinal vertical section of the ship as in the theory of wave resistance for such forms. On the other hand, the keel of the ship may play a large part in wave production in pitching. As a suitable example for calculation, we choose a rectangular form with vertical sides, of length $2l$ and beam $2b$, and floating immersed to a depth f; such a form will clearly exaggerate the wave-making effects of bow and stern. We suppose the form to have an angular pitching oscillation given by $\theta=\theta_0 \sin pt$. We neglect the effect of the vertical ends and consider only the flat base. With the present procedure, we take the source strength at each point such that $4\pi m \cos pt$ is equal to the normal velocity, that is, equal to $x'\dot{\theta}$ in the notation of (22). Hence, from (22), we have in this case

$$\mathrm{P}+i\mathrm{Q}=\frac{p\theta_0}{4\pi}\int_{-l}^{l} dx' \int_{-b}^{b} x' e^{-\kappa_0 f+i\kappa_0(x'\cos\theta+y'\sin\theta)}\,dy' \quad \ldots\ldots \quad (27)$$

$$=\frac{ip\theta_0 \sin(\kappa_0 b \sin\theta)}{\pi\kappa_0^3 \sin\theta\cos^2\theta}\{\sin(\kappa_0 l\cos\theta)-\kappa_0 l\cos\theta\cos(\kappa_0 l\cos\theta)\}e^{-\kappa_0 f}. \quad \ldots\ldots \quad (28)$$

From (23) this gives

$$\mathrm{E}=\frac{8\rho p^3\theta_0^2}{\pi\kappa_0^5}e^{-2\kappa_0 f}\int_0^{\frac{\pi}{2}}\frac{\sin^2(\kappa_0 b\sin\theta)}{\sin^2\theta\cos^4\theta}\{\sin(\kappa_0 l\cos\theta)-\kappa_0 l\cos\theta\cos(\kappa_0 l\cos\theta)\}^2 d\theta. \quad \ldots\ldots \quad (29)$$

For the pitching of a ship, as for heaving, $\kappa_0 b$ is a moderately small quantity; (29) then reduces to a simpler form, which might have been derived directly by assuming a line distribution of sources and sinks. We have then

$$\mathrm{E}=\frac{8\rho p^3 b^2\theta_0^2}{\pi\kappa_0^3}e^{-2\kappa_0 f}\int_0^{\frac{\pi}{2}}\{\sin(\kappa_0 l\cos\theta)-\kappa_0 l\cos\theta\cos(\kappa_0 l\cos\theta)\}^2\frac{d\theta}{\cos^4\theta}. \quad \ldots\ldots \quad (30)$$

For pitching oscillations of a ship, the usual equation for natural pitching in still water is

$$\mathrm{I}\ddot{\theta}+\mathrm{N}\dot{\theta}+g\rho\mathrm{V}m\theta=0, \quad \ldots\ldots \quad (31)$$

where I is the total effective moment of inertia of the ship, V the displaced volume, and m the longitudinal metacentric height. As before, we estimate N by equating the mean value of $\mathrm{N}\dot{\theta}^2$ to the value of E given by (30), with p equal to the natural frequency σ. There do not seem to be any direct determinations or calculations for the added moment of inertia. We shall therefore derive the effective value of I from the relation

$$\sigma^2\mathrm{I}=g\rho\mathrm{V}m, \quad \ldots\ldots \quad (32)$$

with $2\pi/\sigma$ the natural period of pitching when damping is neglected. With this relation, the logarithmic decrement is given by

$$\delta=\frac{\pi N}{\sigma I}=\frac{\pi\sigma N}{g\rho Vm}. \qquad (33)$$

Collecting these results, and expressing the integrand in (30) in terms of Bessel functions, we find

$$\delta=\frac{8\pi\kappa_0 b^2 l^3}{mV}e^{-2\kappa_0 f}\int_0^{\frac{\pi}{2}}J^2_{3/2}(\kappa_0 l\cos\theta)\frac{d\theta}{\cos\theta}. \qquad (34)$$

To obtain a numerical result we take a case from an early paper by Kent (1922). The relevant data for the ship are: length 400 ft., beam 52 ft., draught 22 ft., displacement 10,000 tons, longitudinal metacentric height 458 ft., natural pitching period 6·11 sec.

In (34) we take the distribution of the same length as the ship, that is $l=200$ ft.; we assume a mean beam of 40 ft., and we take $f=20$ ft. These values suffice for a rough approximation. The integral in (34) was computed from tables of Bessel functions; and we obtain finally the result $\delta=1{\cdot}6$. The same general remarks apply to the limitations of this calculation as in the case of heaving motion; however, it is interesting that the decrement δ comes out at about the same value in the two cases.

9. *Summary.*—Using expressions for the wave motion due to alternating sources in a liquid, application is made to the heaving and pitching motions of a ship, and, in particular, to estimating the damping from the rate of propagation of energy outwards in the wave motion. This method of approximation assumes the damping to be small, and the results obtained are too large for much importance to be attributed to the actual numerical values. Nevertheless, it may be concluded that the wave motion gives rise to large damping for both heaving and pitching, and that the decrements are probably comparable with those obtained experimentally.

References.

T. H. Havelock, Phil. Mag. xxix. p. 407 (1940).
F. M. Lewis, Trans. Soc. Nav. Arch. New York, xxxvii. p. 1 (1929).
Browne, Moullin and Perkins, Proc. Camb. Phil. Soc. xxvi. p. 258 (1930).
A. Schuler, *Zeit. f. Ang. Math. u. Mech.* xvi. p. 65 (1936).
F. Horn, *Jahrb. Schiffbau. Ges.* xxxvii. p. 153 (1936).
Kent and Cutland, Trans. Inst. Engrs. and Shipbrs., Scotland, p. 212 (1941).
J. L. Kent, Trans. Inst. Nav. Arch. lxiv. p. 63 (1922).

LXXII. *Relativistic Aspect of the Stress Tensor of the Electromagnetic Field.*

By JESSIE CATTERMOLE, Ph.D.*

[Received May 12, 1942.]

IN a previous paper †, the geodesic equations of the theory of relativity, extended to the five-dimensional form required by the theory of Kaluza, were employed to investigate the motion of a charged particle in a magnetic field, a central electrostatic field, and in a central electrostatic field upon which a uniform magnetic field is superposed. In the present paper it is proposed to investigate the stress tensor of the electromagnetic field of a charged particle from the aspect of Kaluza's theory. Throughout the paper, unless otherwise made clear, Greek indices will indicate five, whereas Latin indices will indicate four possible values, and summation will be implied over repeated indices. Lorentz-Heaviside mixed units will be employed for electric and magnetic quantities, and the field will be supposed to be located in a region of space where the permeability and dielectric constant are each unity.

In the absence of a gravitational field, the components of the Riemann curvature tensor are

$$\gamma_{\alpha\beta}=g_{\alpha\beta}+\frac{K_\alpha K_\beta}{K_5^{\,2}}\gamma_{55}, \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (1)$$

where

$$g_{\alpha\beta}=\begin{cases}0, & \alpha\neq\beta \text{ or } \alpha \text{ or } \beta=5,\\ 1, & \alpha=\beta\neq5.\end{cases}$$

K_1, K_2, K_3 are the components of the electromagnetic vector potential; $K_4=i\frac{V}{c}$, where V is the scalar potential, $i=\sqrt{-1}$, and c is the velocity of light in free space; and K_5 is a constant of suitable dimensions by which electric charge must be multiplied to give the fifth component of momentum.

In the units chosen

$$\left.\begin{aligned}\mathbf{Curl\ K}&=\frac{\mathbf{H}}{c}\\ \frac{\partial V}{\partial x_a}+\frac{\partial K_a}{\partial t}&=-E_a\end{aligned}\right\}\ (a=1\,.\,2\,.\,3), \quad . \; . \; . \; . \; . \; . \quad (2)$$

where **H** and **E** are the magnetic and electric field intensities respectively.

---

* Communicated by the Author.

† Cattermole, Phil. Mag. ser. 7, xxxiii. p. 215 (1942).

It is easy to show that the determinant $|\gamma|$ reduces to γ_{55}, and the contravariant $\gamma^{\mu\nu}$ have the values

$$\gamma^{\mu\nu}=0 \quad (\mu\neq\nu \text{ except } \mu \text{ or } \nu=5),$$

$$\gamma^{\mu\mu}=1 \quad (\text{except } \mu=5),$$

$$\gamma^{5\mu}=\gamma^{\mu 5}=-\frac{\mathrm{K}_\mu}{\mathrm{K}_5}. \quad (\mu\neq5),$$

$$\gamma^{55}=\left(\frac{1}{\lambda_5^{5}}+\frac{\mathrm{K}_1^2+\mathrm{K}_2^2+\mathrm{K}_3^2+\mathrm{K}_4^2}{\mathrm{K}_5^2}\right);$$

Of the 125 Christoffel symbols

$$\Gamma^\sigma_{\mu\nu}=\tfrac{1}{2}\gamma^{\sigma\lambda}\left(\frac{\partial\gamma_{\lambda\mu}}{\partial x^\nu}+\frac{\partial\gamma_{\lambda\nu}}{\partial x^\mu}-\frac{\partial\gamma^{\mu\nu}}{\partial x^\lambda}\right)$$

108 survive, and these are

$$\Gamma^b_{aa}=\gamma_{55}\frac{\mathrm{K}_a}{\mathrm{K}_5^2}\left(\frac{\partial\mathrm{K}_b}{\partial x_a}-\frac{\partial\mathrm{K}_a}{\partial x_b}\right) \text{ not summed,}$$

$$\Gamma^a_{ab}=\frac{\gamma_{55}\mathrm{K}_a}{2\mathrm{K}_5^2}\left(\frac{\partial\mathrm{K}_a}{\partial x_b}-\frac{\partial\mathrm{K}_b}{\partial x_a}\right) \text{ not summed,}$$

$$\Gamma^c_{ab}=\frac{\gamma_{55}}{2\mathrm{K}_5^2}\left\{\mathrm{K}_a\left(\frac{\partial\mathrm{K}_c}{\partial x_b}-\frac{\partial\mathrm{K}_b}{dx_c}\right)+\mathrm{K}_b\left(\frac{\partial\mathrm{K}_c}{\partial x_a}-\frac{\partial\mathrm{K}_a}{\partial x_c}\right)\right\},$$

$$\Gamma^b_{a5}=\frac{\gamma_{55}}{2\mathrm{K}_5}\left(\frac{\partial\mathrm{K}_b}{\partial x_a}-\frac{\partial\mathrm{K}_a}{\partial x_b}\right),$$

$$\Gamma^5_{aa}=\frac{1}{\mathrm{K}_5}\frac{\partial\mathrm{K}_a}{\partial x_a}+\frac{\gamma_{55}\mathrm{K}_a}{\mathrm{K}_5}\left\{\frac{\mathrm{K}_b}{\mathrm{K}_5^2}\left(\frac{\partial\mathrm{K}_a}{\partial x_b}-\frac{\partial\mathrm{K}_b}{\partial x_a}\right)\right\},$$

$$\Gamma^5_{a5}=-\frac{\gamma_{55}}{2\mathrm{K}_5^2}\mathrm{K}_b\left(\frac{\partial\mathrm{K}_b}{\partial x_a}-\frac{\partial\mathrm{K}_a}{\partial x_b}\right),$$

$$\Gamma^5_{ab}=\frac{1}{2\mathrm{K}_5}\left(\frac{\partial\mathrm{K}_a}{\partial x_b}+\frac{\partial\mathrm{K}_b}{\partial x_a}\right)+\frac{\gamma_{55}}{2\mathrm{K}_5^3}\mathrm{K}_c\left\{\mathrm{K}_a\left(\frac{\partial\mathrm{K}_b}{\partial x_c}-\frac{\partial\mathrm{K}_c}{\partial x_b}\right)+\mathrm{K}_b\left(\frac{\partial\mathrm{K}_a}{\partial x_c}-\frac{\partial\mathrm{K}_c}{\partial x_a}\right)\right\}. \quad \ldots\ldots (3)$$

The Riemann-Christoffel tensor $\mathrm{B}^\epsilon_{\mu\nu\sigma}$ of the general relativity theory of the gravitational field, which appears when investigating parallel displacement round a loop, has an exact counterpart in the present theory of the electromagnetic field; and the contracted form, which will be called $\mathrm{R}_{\mu\nu}$, is the counterpart of the $\mathrm{G}_{\mu\nu}$ of gravitational theory.

The $\mathrm{R}_{\mu\nu}$ are given by

$$\mathrm{R}_{\mu\nu}=-\frac{\partial}{\partial x^\sigma}\Gamma^\sigma_{\mu\nu}+\Gamma^\tau_{\mu\sigma}\Gamma^\sigma_{\nu\tau}+\frac{\partial}{\partial x^\mu}\frac{\partial}{\partial x^\nu}\log\sqrt{\gamma|}-\Gamma^\sigma_{\mu\nu}\frac{\partial}{\partial x^\sigma}\log\sqrt{\gamma|}$$

which reduces to

$$\mathrm{R}_{\mu\nu}=-\frac{\partial}{\partial x^\sigma}\Gamma^\sigma_{\mu\nu}+\Gamma^\tau_{\mu\sigma}\Gamma^\sigma_{\nu\tau}, \quad \ldots\ldots\ldots (4)$$

since $|\gamma|=\gamma_{55}=$constant in the present theory.

In what follows α will be written for γ_{55}, and β for K_5.

The $R_{\mu\nu}$, calculated from (4) using (3) and (2), reduce to

$$\left.\begin{aligned}
R_{aa} &= \frac{\alpha}{c\beta^2}\left\{\frac{1}{2c}\left[\mathbf{H}^2-H_a{}^2-E_a{}^2+\frac{\alpha K_a{}^2}{\beta^2}(\mathbf{E}^2-\mathbf{H}^2)\right]\right.\\
&\qquad\qquad \left.-K_a(\mathrm{curl}\ H)_a+\frac{1}{c}K_a\frac{\partial E_a}{\partial t}\right\},\\
R_{ab} = R_{ba} &= \frac{\alpha}{2\beta^2 c}\left\{\frac{1}{c}\left[K_a\frac{\partial E_b}{\partial t}+K_b\frac{\partial E_a}{\partial t}-E_aE_b-H_aH_b\right.\right.\\
&\qquad \left.\left.+\frac{\alpha}{\beta^2}K_aK_b(\mathbf{E}^2-\mathbf{H}^2)\right]-K_a(\mathrm{curl}\ H)_b-K_b(\mathrm{curl}\ H)_a\right\},\\
R_{a4} = R_{4a} &= \frac{i\alpha}{2\beta^2 c}\left\{\frac{1}{c}[\mathbf{E},\ \mathbf{H}]_a-K_a\ \mathbf{div}\ \mathbf{E}+\frac{V}{c^2}\frac{\partial E_a}{\partial t}-\frac{V}{c}(\mathrm{curl}\ H)_a\right.\\
&\qquad\qquad \left.+\frac{\alpha K_a V}{\beta^2 c^2}(\mathbf{E}^2-\mathbf{H}^2)\right\},\\
R_{44} &= \frac{\alpha}{\beta^2 c^2}\left\{V\ \mathbf{div}\ \mathbf{E}-\tfrac{1}{2}\mathbf{E}^2-\frac{\alpha V^2}{2\beta^2 c^2}(\mathbf{E}^2-\mathbf{H}^2)\right\},\\
R_{a5} = R_{5a} &= \frac{\alpha}{2\beta c}\left\{\frac{1}{c}\frac{\partial E_a}{\partial t}-(\mathrm{curl}\ H)_a+\frac{\alpha K_a}{\beta^2 c}(\mathbf{E}^2-\mathbf{H}^2)\right\},\\
R_{45} = R_{54} &= \frac{i\alpha}{2\beta c}\left\{\frac{\alpha V}{\beta^2 c^2}(\mathbf{E}^2-\mathbf{H}^2)-\mathbf{div}\ \mathbf{E}\right\},\\
R_{55} &= \frac{\alpha^2}{2\beta^2 c^2}\{\mathbf{E}^2-\mathbf{H}^2\}.
\end{aligned}\right\} \quad (5)$$

In (5) and the rest of the paper, the Latin suffices have values 1, 2, 3. The components of the mixed tensor $R_\nu{}^\sigma=\gamma^{\sigma\mu}R_{\mu\nu}$ are

$$\left.\begin{aligned}
R_a{}^a &= \frac{\alpha}{2\beta^2 c}\left\{\frac{1}{c}\left(K_a\frac{\partial E_a}{\partial t}-E_a{}^2+\mathbf{H}^2-H_a{}^2\right)-K_a(\mathrm{curl}\ H)_a\right\},\\
R_a{}^b = R_b{}^a &= \frac{\alpha}{2\beta^2 c}\left\{\frac{1}{c}\left(K_a\frac{\partial E_b}{\partial t}-E_aE_b-H_aH_b\right)-K_a(\mathrm{curl}\ H)_b\right\},\\
R_4{}^a = R_a{}^4 &= \frac{i\alpha}{2\beta^2 c^2}\left\{\frac{V}{c}\frac{\partial E_a}{\partial t}+[\mathbf{E},\ \mathbf{H}]_a-V(\mathrm{curl}\ H)_a\right\},\\
R_4{}^4 &= \frac{\alpha}{2\beta^2 c^2}\{V\ \mathbf{div}\ \mathbf{E}-\mathbf{E}^2\},\\
R_5{}^a &= \frac{\alpha}{2\beta c}\left\{\frac{1}{c}\frac{\partial E_a}{\partial t}-(\mathrm{curl}\ H)_a\right\},\\
R_a{}^5 &= \frac{1}{2\beta c}\left\{\frac{1}{c}\frac{\partial E_a}{\partial t}-(\mathrm{curl}\ H)_a\right\}\\
&\quad +\frac{\alpha}{2\beta^3 c}\left\{\frac{V}{c^2}[\mathbf{E},\ \mathbf{H}]+\frac{1}{c}[H_a(\mathbf{K},\ \mathbf{H})+E_a(\mathbf{K},\ \mathbf{E})]\right.\\
&\quad \left.+K_a\left[(\mathbf{K},\ \mathbf{curl}\ \mathbf{H})-\frac{1}{c}\left(\mathbf{K},\frac{\partial\mathbf{E}}{\partial t}\right)-\frac{V}{c}\ \mathbf{div}\ \mathbf{E}+\frac{1}{c}(\mathbf{E}^2-2\mathbf{H}^2)\right]\right\},\\
R_5{}^4 &= \frac{\alpha}{2\beta ic}\ \mathbf{div}\ \mathbf{E},
\end{aligned}\right\} \quad (6)$$

$$\left.\begin{aligned}
\mathrm{R}_4{}^5 &= \frac{1}{2\beta ic}\,\mathrm{div}\,\mathbf{E} + \frac{i\alpha}{2\beta^3c^2}\left\{\frac{\mathrm{V}}{c}(2\mathbf{E}^2-\mathbf{H}^2)+\mathrm{V}\big(\mathbf{K},\,\mathrm{curl}\,\mathbf{H}\big)\right.\\
&\qquad\left.-\frac{\mathrm{V}}{c}\left(\mathbf{K},\frac{\partial\mathbf{E}}{\partial t}\right)-\big(\mathbf{K},\,[\mathbf{E},\,\mathbf{H}]\big)-\frac{\mathrm{V}^2}{c}\,\mathrm{div}\,\mathbf{E}\right\},\\
\mathrm{R}_5{}^5 &= \frac{\alpha}{2\beta^2c}\left\{-\frac{\mathbf{H}^2}{c}+\frac{\mathbf{E}^2}{c}-\frac{\mathrm{V}}{c}\,\mathrm{div}\,\mathbf{E}-\frac{1}{c}\left(\mathbf{K},\frac{\partial\mathbf{E}}{\partial t}\right)+\big(\mathbf{K},\,\mathrm{curl}\,\mathbf{H}\big)\right\}.
\end{aligned}\right\}\quad . \qquad \tag*{(6) (cont.)}$$

The spur $\mathrm{R}=\gamma^{\mu\nu}\mathrm{R}_{\mu\nu}=\gamma_\sigma{}^\nu\mathrm{R}_\nu{}^\sigma,$

where $\gamma_\sigma{}^\nu=\begin{cases}1, & \sigma=\nu,\\ 0, & \sigma\neq\nu,\end{cases}$

has the value
$$\mathrm{R}=\frac{\alpha}{2\beta^2c^2}(\mathbf{H}^2-\mathbf{E}^2), \quad . \; . \; . \; . \; . \; . \; . \; . \tag{7}$$

so that the $\mathrm{R}_\nu{}^\sigma-\frac{1}{2}\gamma_\nu{}^\sigma\mathrm{R}$ are

$$\left.\begin{aligned}
\mathrm{R}_a{}^a-\tfrac{1}{2}\mathrm{R} &= \frac{\alpha}{2\beta^2c}\left\{\frac{1}{c}\left(\mathrm{K}_a\frac{\partial\mathrm{E}_a}{\partial t}-\mathrm{E}_a{}^2+\tfrac{1}{2}\mathbf{E}^2\right)-\mathrm{K}_a\,(\mathrm{curl}\,\mathrm{H})_a-\frac{\mathrm{H}_a{}^2}{c}+\frac{I}{2}\frac{\mathbf{H}^2}{c}\right\},\\
\mathrm{R}_4{}^4-\tfrac{1}{2}\mathrm{R} &= \frac{\alpha}{2\beta^2c^2}\{\mathrm{V}\,\mathrm{div}\,\mathbf{E}-\tfrac{1}{2}\mathbf{E}^2-\tfrac{1}{2}\mathbf{H}^2\},\\
\mathrm{R}_5{}^5-\tfrac{1}{2}\mathrm{R} &= \frac{\alpha}{2\beta^2c}\left\{\big(\mathbf{K},\,\mathrm{curl}\,\mathbf{H}\big)-\frac{1}{c}\left(\mathbf{K},\frac{\partial\mathbf{E}}{\partial t}\right)-\frac{\mathrm{V}}{c}\,\mathrm{div}\,\mathbf{E}+\frac{3}{2c}(\mathbf{E}^2-\mathbf{H}^2)\right\},\\
\text{and}\quad & \mathrm{R}_a{}^b=\mathrm{R}_b{}^a\,;\ \mathrm{R}_4{}^a=\mathrm{R}_a{}^4\,;\ \mathrm{R}_5{}^a\,;\ \mathrm{R}_a{}^5\,;\ \mathrm{R}_5{}^4\,;\ \mathrm{R}_4{}^5,\\
&\text{which have the values given in (6).}
\end{aligned}\right\}\quad . \; . \; . \tag{8}$$

If the particle be uncharged—a photon for example—

$$\left.\begin{aligned}
\mathrm{curl}\,\mathbf{H} &= \frac{1}{c}\frac{\partial\mathbf{E}}{\partial t},\\
\mathrm{div}\,\mathbf{E} &= 0,
\end{aligned}\right\}\quad . \; . \; . \; . \; . \; . \; . \; . \tag{9}$$

and the components of (8) reduce to

$$\left.\begin{aligned}
&\mathrm{R}_a{}^a-\tfrac{1}{2}\mathrm{R}=-\frac{\alpha}{2\beta^2c^2}\{\mathrm{H}_a-\tfrac{1}{2}\mathbf{H}^2+\mathrm{E}_a{}^2-\tfrac{1}{2}\mathbf{E}^2\},\\
&\mathrm{R}_a{}^b=\mathrm{R}_b{}^a=-\frac{\alpha}{2\beta^2c^2}\{\mathrm{H}_a\mathrm{H}_b+\mathrm{E}_a\mathrm{E}_b\},\\
&\mathrm{R}_4{}^a=\mathrm{R}_a{}^4=-\frac{\alpha}{2\beta^2c^2}\{-i[\mathbf{E},\,\mathbf{H}]_a\},\\
&\mathrm{R}_4{}^4-\tfrac{1}{2}\mathrm{R}=-\frac{\alpha}{2\beta^2c^2}\{\tfrac{1}{2}\mathbf{H}^2+\tfrac{1}{2}\mathbf{E}^2\},\\
&\mathrm{R}_5{}^a=0,\\
&\mathrm{R}_a{}^5=\frac{\alpha}{2\beta^3c^2}\left\{\frac{\mathrm{V}}{c}[\mathbf{E},\,\mathbf{H}]_a+\mathrm{E}_a\big(\mathbf{K},\,\mathbf{E}\big)+\mathrm{K}_a(\mathbf{E}^2-2\mathbf{H}^2)\right\},\\
&\mathrm{R}_5{}^4=0,\\
&\mathrm{R}_4{}^5=\frac{i\alpha}{2\beta^3c^3}\left\{\frac{\mathrm{V}}{c}(2\mathbf{E}^2-\mathbf{H}^2)-\big(\mathbf{K},\,[\mathbf{E},\,\mathbf{H}]\big)\right\},\\
&\mathrm{R}_5{}^5-\tfrac{1}{2}\mathrm{R}=-\frac{\alpha}{2\beta^2c^2}\{\tfrac{3}{2}\mathbf{H}^2-\tfrac{3}{2}\mathbf{E}^2\}=0\\
&\text{for a sinusoidal photon wave, since } \mathbf{H}=\mathbf{E}.
\end{aligned}\right\}\quad . \; . \tag{10}$$

Ignoring R_a^5 and R_4^5,* (10) taken with a negative sign are the components of the electromagnetic stress tensor for a photon field if we set

$$\frac{\alpha}{2\beta^2c^2}=1, \qquad (11)$$

as we are adopting Lorentz-Heaviside mixed units. We are thus led to conclude that (8) are the components of the stress tensor for an electron (or positron) field.

For such a field we have †

$$\mathbf{div}\ \mathbf{E}=\rho,$$

where ρ is the charge density, and

$$R_4^4-\tfrac{1}{2}R=-\frac{\alpha}{2\beta^2c^2}\{-V\rho+\tfrac{1}{2}\mathbf{E}^2+\tfrac{1}{2}\mathbf{H}^2\}.$$

$V\rho$ is the work or energy per unit volume of the field of the charged particle, and it must be added to (or subtracted from) that of the photon field, according to the sign of ρ.

From (8) also, since $\mathbf{curl}\ \mathbf{H}-\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}=\frac{\rho v'}{c}$, where v' is the velocity of transport of charge, we have

$$R_4^a=\frac{i\alpha}{2\beta^2c^2}\left\{[\mathbf{E},\mathbf{H}]_a-\frac{V\rho v'}{c}\right\},$$

so that Poynting's vector no longer represents the total of the density of flow of electromagnetic energy.

From (11),

$$\alpha=2c^2\beta^2$$

or

$$\gamma_{55}=2c^2K_5^2,$$

and, since in the present theory $eK_5=m_0c$, where e and m_0 are the charge and "rest" mass of the particle,

$$\gamma_{55}=\frac{2m_0^2c^4}{e^2},$$

which is the value of the hitherto undetermined constant γ_{55}.

---

* The tensor $R_\mu^\nu-\frac{1}{2}\gamma_\mu R$ does not appear to be symmetrical, as it undoubtedly should be; but the author believes that the asymmetry is merely apparent and that restoration of symmetry should not present any great difficulty. The way in which this is achieved, however, is not clear at the time of writing.

† Wilson and Cattermole, Phil. Mag. ser. 7, xxvii. p. 84 (1939).

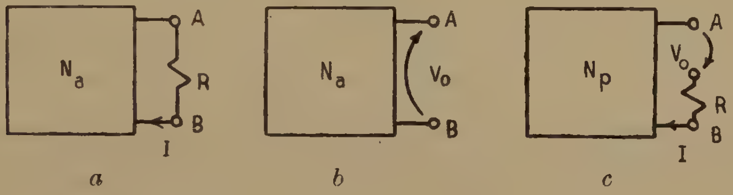
LXXIII. *A General Superposition Theorem of the “Thevenin” Type.*

By G. F. **Freeman**, M.Sc.(Eng.), A.M.I.E.E., A.M.I.Mech.E.,
West Ham Municipal College *.

[Received July 1, 1942.]

Interest has recently been renewed in the superposition device usually attributed to Thevenin (1883), but as Wall has pointed out †, anticipated by Helmholtz (1853) and later extended by Pleijel to alternating conditions. Briefly, the method consists in finding the current I flowing in a resistance branch R between terminals A and B (fig. 1 *a*) of an active d.c. network N_a (*i. e.* a network containing one or more e.m.f. sources) by removing R, obtaining the p.d. V_0 across the open circuit so formed (fig. 1 *b*), and applying this voltage to R in series with the network, now rendered passive by the replacement of each e.m.f. source by a short circuit, and styled N_p; the resultant current (fig. 1 *c*) in R is I. In the particular case of the determination of rotor current in the equivalent circuit of a polyphase induction motor, the corresponding a.c. process goes by the name of Ho's theorem ‡. Phase as well as magnitude is preserved. Writing Z for R, and Z_n for the impedance of N_p between A and B, the result may be stated as

Fig. 1.



$$I=V_0/(Z+Z_n), \qquad (1)$$

which also includes the d.c. case.

It is proposed to show that a more general theorem exists, and to instance some of its possibilities by selected examples. The relationships are vector; d.c. conditions may be regarded as a special case in which all vectors are collinear, but may of course have positive or negative sign.

In fig. 2 *a* is represented an active network N_a, across two terminals A and B, of which is connected an impedance Z. The impelling p.d.

---

* Communicated by the Author.

† ‘Electrical Review,’ p. 501 (Nov. 7, 1941).

‡ Karapetoff ‘Electrical Engineering’ (later editions).

V across Z sends a current I through it. The reciprocal of Z is its admittance Y. Fig. 2 *b* shows similar conditions for some other impedance Z′, the other quantities being V′, I′ and Y′ respectively. Then it is shown in Appendix I. that

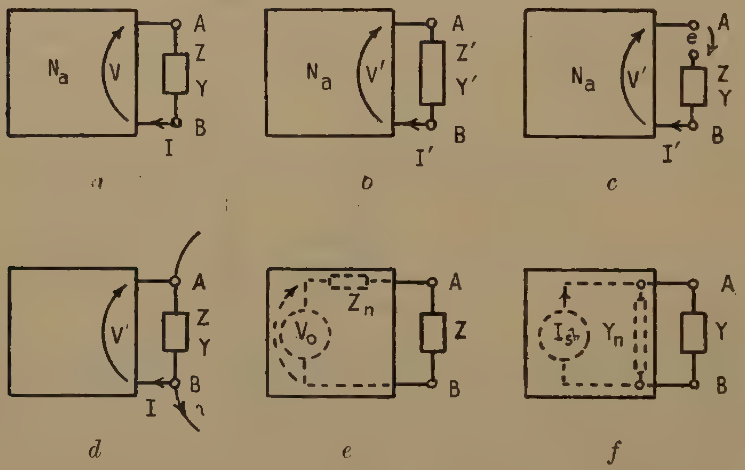
$$I'(Z'+Z_n)=I(Z+Z_n)=V_0 \quad \ldots\ldots \quad (2)$$

and

$$V'(Y'+Y_n)=V(Y+Y_n)=I_s. \quad \ldots\ldots \quad (2\,a)$$

These two complementary relations, either of which may be obtained from the other by multiplication or division by Z_n or its reciprocal Y_n, are believed to be the most general and symmetrical forms in which relationships of this type may be described. They lead to two alternative and equivalent conceptions of the network, which may be treated either as a source of *constant e.m.f.* V_0 (fig. 2 *e*) in *series* with an *impedance* Z_n or as a source of *constant current* I_s (fig. 2 *f*) in *shunt* with an *admittance* Y_n. Either representation may be used, according to circumstances. Obviously, V_0 is the open circuit voltage, and I_s the short-circuit current, at AB.

Fig. 2.



We have also, directly,

$$V'+I'Z_n=V+IZ_n, \quad \ldots\ldots \quad (3)$$

$$I'+V'Y_n=I+VY_n. \quad \ldots\ldots \quad (3\,a)$$

a useful, but less symmetrical, way of stating the relations.

Open and short circuit conditions give the special cases

$$I=V_0/(Z+Z_n) \text{ (already noted)}, \quad \ldots\ldots \quad (1)$$

$$V=I_s/(Y+Y_n), \quad \ldots\ldots \quad (1\,a)$$

$$I=I_sZ_n/(Z+Z_n), \quad \ldots\ldots \quad (4)$$

$$V=V_0Y_n/(Y+Y_n). \quad \ldots\ldots \quad (4\,a)$$

The familiar and valuable reciprocal relations

$$V_0 = I_s Z_n, \qquad (5)$$

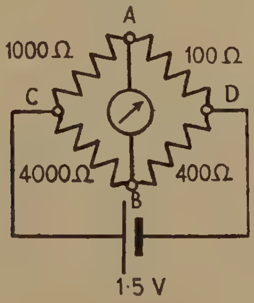
$$I_s = V_0 Y_n \qquad (5\,a)$$

are obvious corollaries.

Since constant voltage, rather than constant current, is the prevailing condition of electrical supply, it might appear that constant current methods are of less practical interest. However, it often happens that a solution sufficiently exact for ordinary purposes may be obtained in what is strictly a constant voltage problem by ignoring minor changes in current supply, with considerable saving in time; it all depends on the relative magnitudes involved. The extent to which such a technique may be used is largely a matter of experience.

The examples which follow have been chosen to illustrate certain aspects of the methods, and are drawn from familiar backgrounds in order to permit of easy verification by other means. No attempt has been made at this stage to display the processes to their greatest advantage; common sense and a little practice will soon indicate where they may best be used, once acquaintance and confidence have been established.

Fig. 3.



Example A.—Galvanometer Sensitivity in Four-Arm Bridge Network.

For use in the Wheatstone net of fig. 3, it is premised that two galvanometers are available: (*a*) resistance 1000Ω, deflecting 17 cm. per μA (0·017 cm. per μV) at 1 m.; (*b*) resistance 40Ω, deflecting 1·5 cm. per μA (0·0375 cm. per μV) at 1 m. It is required to estimate their relative indications as detectors in the circuit, and then to deduce the sensitivity of the bridge at balance, using the preferred galvanometer.

Here the galvanometer is "Z," and "Z_n," the network impedance between A and B is 455Ω, by the ordinary rules of resistance combination, when C and D are short-circuited to achieve the passive state.

Assume a current I_a to flow in AB, using galvanometer (*a*), so small that the value of Z_n is sensibly the same as at balance. We may then write with negligible error, by (2),

$$\frac{I_b}{I_a} = \frac{1000+455}{40+455} = 2{\cdot}94,$$

where I_b is the corresponding current in galvanometer (b) for the same unbalance of the arms. The ratio of deflexions, using the current sensitivities, is thus

$$\frac{d_a}{d_b}=\frac{17\,I_a}{1\cdot5\,I_b}=\frac{17}{1\cdot5\times2\cdot94}=3\cdot85.$$

Alternatively, $Y_n=1/Z_n=0\cdot00220$ mho, $Y_a=0\cdot001$ mho, $Y_b=0\cdot025$ mho. Thence, by (2 a),

$$\frac{V_b}{V_a}=\frac{0\cdot001+0\cdot00220}{0\cdot025+0\cdot00220}=0\cdot118,$$

whence, using in this case the voltage sensitivities,

$$\frac{d_a}{d_b}=\frac{0\cdot017\,V_a}{0\cdot0375\,V_b}=\frac{0\cdot017}{0\cdot0375\times0\cdot118}=3\cdot85$$

as before.

To find the sensitivity of the bridge it is convenient to estimate the unbalance of the arms for a selected small galvanometer current, not enough to affect the value of Z_n. Using galvanometer (a), we have for $I_a=1\times10^{-8}$ A, to an excellent approximation, by (2),

$$V_0=(1000+455)\times10^{-8}=14\cdot55\ \mu V.$$

Assuming the tapping point to have moved to the left along the arm AC to effect this unbalance, we have also, if x is the percentage departure,

$$V_0=1\cdot5\left(\frac{100+x}{1100}-\frac{400}{4400}\right)\text{ volts.}$$

Eliminating V_0, $x=0\cdot0107$ per cent.; a deflexion of 1·7 mm. at 1 m. thus corresponds to an unbalance of 1 in 10,000.

Example B.—Short-circuit Currents and Regulation in Power Network.

The possible line-to-earth current for a total short-circuit at a load terminal on a three-phase line (and hence the required rupturing capacity for a circuit breaker installed there) is to be estimated, given that the line voltage is $132\overline{\diagdown 0^\circ}$ kV on open circuit and $120\overline{\diagdown 5^\circ}$ kV when supplying a balanced load of $50\overline{\diagdown 30^\circ}$ MVA.

In terms of one star phase,

$$Z'=\frac{120^2\times10^6}{50\overline{\diagdown 30^\circ}\times10^6}=288\underline{/30^\circ}\ \Omega\ ;\quad V'=69\cdot3\overline{\diagdown 5^\circ}\ kV\ ;\quad I'=0\cdot241\overline{\diagdown 35^\circ}\,kA.$$

$$\text{From (2),}\quad Z_n=(V_0/I')-Z'=\frac{76\cdot2\overline{\diagdown 0^\circ}}{0\cdot241\overline{\diagdown 35^\circ}}-288\underline{/30^\circ}=38\cdot4\underline{/75^\circ}\ \Omega.$$

$$\text{Thence by (5),}\quad I_s=(V_0/Z_n)=\frac{76\cdot2\overline{\diagdown 0^\circ}}{38\cdot4\underline{/75^\circ}}=1\cdot99\overline{\diagdown 75^\circ}\ kA.$$

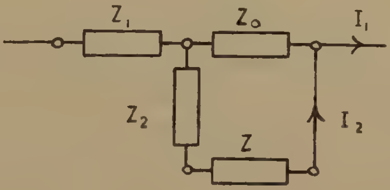
The rupturing capacity must then be not less than $3\times76\cdot2\times1\cdot99=455$ MVA., assuming that all connected plant behaves as if e.m.f.'s and impedances are constants. (The degree to which transient asymmetry, magnetic saturation and automatic voltage regulators affect the result would have to be taken into account, but in many cases this figure would suffice.)

Problems involving regulation at one load when it is known at another may be handled in the same way. A percentage basis is convenient, and unbalanced loadings may be treated by symmetrical components, bearing in mind that impedance values will be particular to the component concerned—it is not generally possible to predict regulation on negative or zero component from data exclusive to positive component (balanced loading). It is an advantage if the known loading is relatively heavy, since the determination of Z_n as above depends on differences, and phase displacement between V and V_0 is not easily measured in the field, although perfectly possible to determine on a model. If, of course, Z_n is known, regulation on any load may be surmised with great facility, while the absence of phase effect makes the corresponding d.c. problems very easy to handle. Estimation of Z_n on a purely reactance basis is a useful approximate device.

Example C.—Current Transformer Performance.

Using the approximate equivalent circuit of fig. 4, the influence of magnetizing impedance Z_0 and secondary leakage impedance Z_2 upon the performance of a current transformer at various burdens Z may be investigated. (Primary impedance Z_1 is not significant in view of the constant current source.) Denoting the impedances in terms of their admittance reciprocals, the short circuit condition gives

Fig. 4.



$$I_s = I_1 Y_2 (Y_0 + Y_2),$$

whence by (1 *a*) the burden voltage

$$V = \frac{I_s}{Y + \dfrac{Y_0 Y_2}{Y_0 + Y_2}} = \frac{I_s (Y_0 + Y_2)}{YY_0 + YY_2 + Y_0 Y_2} = \frac{I_1 Y_2}{YY_0 + YY_2 + Y_0 Y_2}.$$

The fractional error

$$E = \frac{I_1 - I_2}{I_1} = 1 - \frac{VY}{I_1}$$

$$= \frac{1}{1 + \dfrac{Z_0}{Z + Z_2}}.$$

Both currents are of course referred to secondary turns.

Conclusion.

It is hoped that these few examples may at any rate indicate some of the possible lines of development, and suggest the wide variety of subjects to which such methods lend themselves. They may, in fact, be introduced whenever they add speed to a solution without sacrificing accuracy and intelligibility.

In combination (not competition) with star-mesh transformation, devised by Kennelly in 1899, many more complicated problems—*e. g.* six-arm a.c. bridge details—may be readily resolved. Anyone sufficiently interested to experiment with the methods will quickly discover the powerful and profitable aid they afford in what tends at times to be a very tedious business—the calculation of networks, especially a.c. networks.

APPENDIX I.

Consider two terminals A and B of an active network N_a to be closed (fig. 2 *a*) by an impedance Z, which is replaced in fig. 2 *b* by another impedance Z′. Let the voltages and currents be V, I, V′ and I′ respectively, and write $Y=(1/Z)$, $Y'=(1/Z')$. It is required to show that

$$I'(Z'+Z_n)=I(Z+Z_n)=V_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

and that
$$V'(Y'+Y_n)=V(Y+Y_n)=I_s, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2\,a)$$

where Z_n is the passive impedance of the network between A and B and $Y_n=(1/Z_n)$ the corresponding admittance.

First Proof.

Introduce (fig. 2 *c*) an e.m.f. *e* in series with the circuit of fig. 2 *a*, such that the voltage and current assume the values V′ and I′ appropriate to fig. 2 *b*. Then

$$V'+e=I'Z.$$

But the current change is by superposition

$$I'-I=e/(Z+Z_n).$$

Eliminating *e* from these two expressions and writing $V'=I'Z'$,

$$I'(Z'+Z_n)=I(Z+Z_n)=V_0 \text{ (say)}. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The relation (2 *a*) follows directly if both sides are multiplied by Y_n. The constant V_0 is clearly the terminal voltage when $Z\rightarrow\infty$, and thus is the e.m.f. of an equivalent constant voltage source.

Second Proof.

Introduce (fig. 2 *d*) a current *i* in shunt with the circuit of fig. 2 *a*, such that the voltage and current assume the values V′ and I′ appropriate to fig. 2 *b*. Then

$$I'+i=V'Y.$$

But the voltage change is by superposition

$$V'-V=i/(Y+Y_n).$$

Eliminating i from these two expressions and writing $I'=V'Y'$,

$$V'(Y'+Y_n)=V(Y+Y_n)=I_s \text{ (say)}, \quad . \quad . \quad . \quad . \quad (2\,a)$$

from which again (2) follows if both sides are multiplied by Z_n. The constant I_s is evidently the short-circuit current when $Y\rightarrow\infty$, and leads to the conception of a constant current source.

The two representations are shown diagrammatically in fig. 2 *e* and fig. 2 *f*, in which the characteristic equivalent sources V_0 and I_s are evident.

LXXIV. *The Viscosity-Temperature Function of Liquids.*

By E. P. IRANY *.

[Received May 26, 1942.]

UNDER the above title, A. H. Nissan (1) published an article in a recent issue of this Journal in which he demonstrated the existence of a general relation between liquid viscosity and temperature. He employed a graphical method resting on the assumption that the viscosity-temperature function of the normal paraffins approaches the fundamental law closely enough to serve as a standard of comparison for all other liquids; he introduced a system of co-ordinates which makes the strongly curved plottings of viscosity-temperature functions appear linear; he showed that if the functional scale derived from one normal paraffin is applied to other members of the series, all are represented on straight lines; that the same holds true for " all pure liquids provided they do not change identity . . . . with temperature change," and that " the viscosity of any liquid can be calculated, using the chart, for wide variations in temperature from only two experimental values."

The principle of the functional scale on which Nissan's method of graphical analysis is based, as well as all deductions enumerated above, have been disclosed by myself more than four years preceding the appearance of Nissan's article (2). While scant recognition had been paid to this fact in the form of a footnote in an earlier paper by Nissan, Clark and Nash (3) on the same subject, I regret to note that in the paper under discussion even this acknowledgment has been omitted.

I certainly do not wish to detract from but appreciate the large amount of data with which Nissan exemplifies his work and, therewith, confirms my own postulates. However, the method of graphical representation which he applies is only in appearance different from the one used by me. It is easy to prove that, actually, his treatment of the problem is identical with mine and cannot but lead to the same conclusions.

---

* Communicated by the Author.

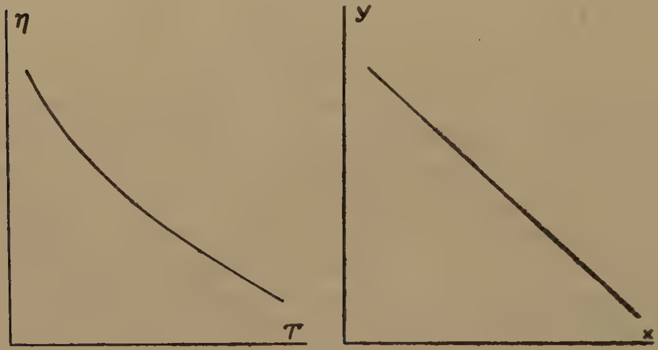
Let the natural function between absolute viscosity, η, and absolute temperature, T, for a normally behaving (ideal) liquid be represented by a curve in the η—T diagram as in fig. 1. It may then be assumed that a system of two co-ordinates $x=F(T)$ and $y=G(\eta)$ exists in which this curve would appear linear,

$$\frac{dy}{dx}=m=\text{constant}. \qquad (1)$$

Hence, as a general requirement for x and y,

$$\frac{dy}{dx}=\frac{d\eta}{dT}\frac{\frac{dy}{d\eta}}{\frac{dx}{dT}}=\frac{d\eta}{dT}\frac{G'}{F'}=m. \qquad (2)$$

Fig. 1.



In my original work I made

$$X=F(T)=T, \quad y=G(\eta)=\theta,$$

$$F'=1, \quad G'=\frac{d\theta}{d\eta}, \quad \frac{d\eta}{dT}\frac{d\theta}{d\eta}=m_1. \qquad (3)$$

Combining (2) and (3), the relation between other employable scale functions is given by

$$\frac{G'}{F'}=\frac{m}{m_1}\frac{d\theta}{d\eta}. \qquad (4)$$

For example, in Nissan's case,

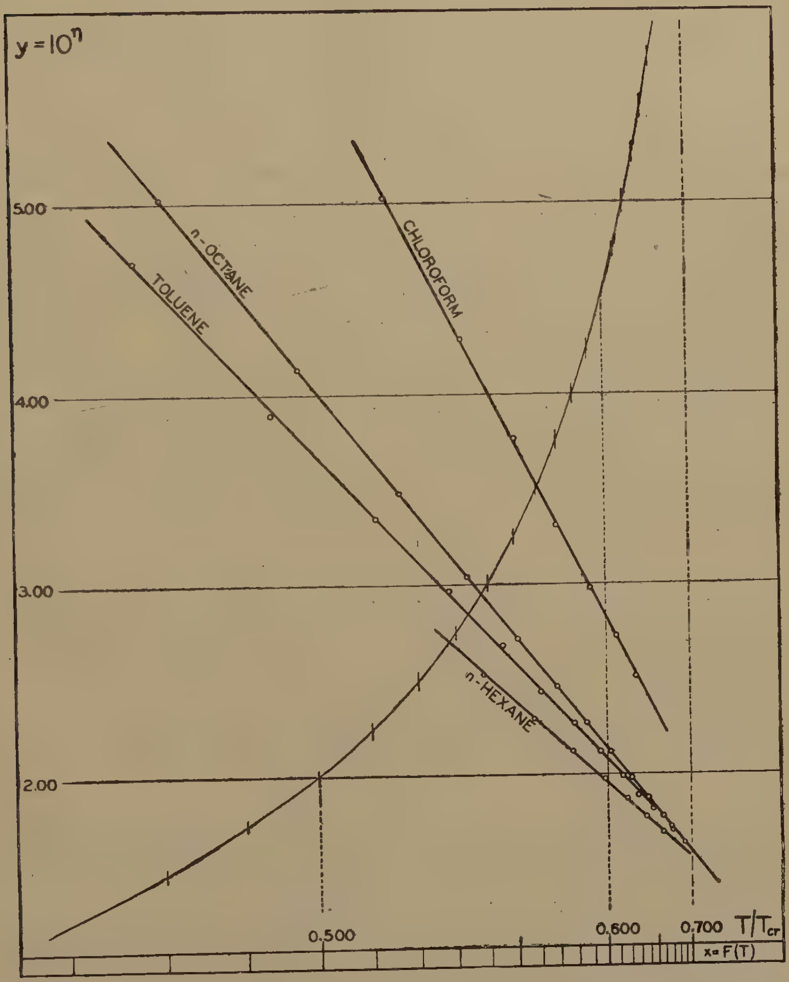
$$X=F(T), \quad y=G(\eta)=\log\eta,$$

$$F'=\frac{m_1}{m}\log e\frac{1}{\eta\frac{d\theta}{d\eta}}=\frac{K}{\eta\frac{d\theta}{d\eta}}, \qquad (5)$$

showing his temperature scale F(T) entirely determined by absolute viscosity as such and my viscosity function θ. It has been the object of my prior work to prove that the latter is a unique function of viscosity, independent of the nature of any particular liquid under survey and independent of the temperature level at which the given viscosity is attained. According to equation (5), Nissan's temperature scale F(T) is fully determined by viscosity itself and the function $\theta(\eta)$; hence, it is nothing more than a restatement of the latter in implicit form.

Equation (4) expresses the fact that for any arbitrarily assumed continuous function of viscosity a temperature scale can be found, and

Fig. 2.



for every continuous temperature function a viscosity scale, which satisfies the provision of linearity. For example, let Nissan's preconceived viscosity function $y=\log \eta$ be reversed into the entirely meaningless function $y=10^{\eta}$. Choosing *n*-octane as a standard substance, a temperature scale can be constructed which again straightens all viscosity-temperature plottings, as shown for *n*-hexane, toluene and chloroform in fig. 2, using Thorpe and Rodger's data (4). The same result was obtained with other equally empty formulations such as η^2, $\sqrt[3]{\eta}$, arc tan η, $\sqrt{T/T_{cr}}$. However, this does not mean that all these combinations of co-ordinates are equally useful for interpretation. Temperature—whether expressed in absolute, relative or reduced units—is the only quantity involved here which possesses assured and fundamental additivity. Consequently, the directly correlated viscosity function θ assumes unique significance.

As I explained in a later paper on the subject (5), the functions θ and log η become nearly indistinguishable at higher viscosities, so that Nissan's temperature scale should approach constancy of increments. However, in the lower viscosity levels the deviations are very large.

The mathematical formulation of Nissan's temperature function relating log η and T uses the conventional approximation by a power series which is applicable to any function if an adequate number of constants are provided. In Nissan's case inclusion of the cubical term, and hence four constants, accurately reproduce three-place values. As an empirical means for interpolation and approximation this formula is quite legitimate, but it certainly must be kept distinct from a rigorously derived parabolic equation of the third order. Nissan's deductions by far transcend the origin of his formula. A negative sign in an algebraic sum of this kind hardly warrants far-reaching physical interpretation. The underlying precept of the sameness of the liquid and gaseous states under the common term of "fluids" is a tempting generalization, but it has done more harm than good to our understanding of the nature of liquids (6).

Literature References.

(1) A. H. Nissan, Phil. Mag. xxxii. pp. 441–56 (1941).
(2) E. P. Irany, Journ. Am. Chem. Soc. lx. pp. 2106–15 (1938).
(3) A. H. Nissan, L. V. W. Clark and A. W. Nash, Journ. Inst. Petrol. xxvi. pp. 155–222 (1940).
(4) Thorpe and Rodger, Phil. Trans. clxxxv. p. 397 (1896).
(5) E. P. Irany, J. Amer. Chem. Soc. lxiii. pp. 2611–17 (1941).
(6) E. P. Irany, *ibid.* lxi. pp. 1436–43 (1939).

LXXV. *Approximate Expressions for Transmission Line Constants in Terms of the Open- and Closed-circuit Impedances.*

By J. C. SIMMONDS *.

[Received June 2, 1942.]

It can be shown that the primary constants of a transmission line are related to the open- and closed-circuit impedances of the line by the following equations :—

$$R+j\omega L=PZ_0, \quad \ldots \quad (1)$$

---

* Communicated by the Author.

$$G+j\omega C=P/Z_0, \quad \ldots \quad (2)$$

$$Pl=\tanh^{-1}\sqrt{\frac{Z_{sc}}{Z_{oc}}}, \quad \ldots \quad (3)$$

$$Z_0=\sqrt{Z_{sc}\,.\,Z_{oc}}, \quad \ldots \quad (4)$$

where R is the resistance of the line per unit length,
G is the leakance of the line per unit length,
L is the inductance of the line per unit length,
C is the capacitance of the line per unit length,
P is the propagation constant of the line per unit length,
Z_0 is the characteristic impedance of the line,
Z_{sc} is the input impedance of the line with the far end closed,
Z_{oc} is the input impedance of the line with the far end opened,
l is the length of the line.

From these equations the primary constants (*i. e.* the resistance, leakance, inductance and capacitance per unit length of line), the characteristic impedance and the propagation constant, as well as certain other related quantities, can be calculated once the open- and closed-circuit impedances of the line have been measured. However, in general both the open- and the closed-circuit impedances will be complex quantities, and therefore considerable manipulation will be necessary before the required results are obtained. Further, it is not easy to ascertain the number of figures to which the calculations must be carried to obtain a given accuracy in the final result. When transmission lines are used at radio frequencies certain approximations can be made and far simpler, though approximate, expressions derived for the characteristic impedance and for the primary constants, as will now be shown.

Let the open- and closed-circuit impedances of the transmission line be

$$Z_{oc}=\frac{1}{G_m+j\omega C_m}, \quad \ldots \quad (5)$$

and

$$Z_{sc}=R_m+j\omega L_m \quad \ldots \quad (6)$$

the " *m* " suffixes denoting measured quantities.

Characteristic Impedance.

From equations (4), (5) and (6) the characteristic impedance can be expressed in the form

$$Z_0=\sqrt{\frac{L_m}{C_m}\left\{\left(1-\frac{jR_m}{\omega L_m}\right)\left(1-\frac{jG_m}{\omega C_m}\right)^{-1}\right\}}, \quad \ldots \quad (7)$$

and after some manipulation this can be shown to be

$$Z_0=\sqrt{\frac{L_m}{C_m}}\left\{1-\frac{j}{2}\left(\frac{R_m}{\omega L_m}-\frac{G_m}{\omega C_m}\right)\right\} \quad \ldots \quad (8)$$

to within 1 per cent. if $R_m/\omega L_m$ and $G_m/\omega C_m$ are not greater than 0·1, and to within 0·1 per cent. if they are not greater than 0·04.

Resistance and Inductance.

From equations (1), (3) and (4) the resistance and the inductance are seen to be given by

$$\mathrm{R}+j\omega\mathrm{L}=\frac{\mathrm{Z}_{sc}}{l}\sum_{n=1}^{\infty}\frac{1}{(2n-1)}\cdot\left(\frac{\mathrm{Z}_{sc}}{\mathrm{Z}_{oc}}\right)^{n-1}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

providing $|\mathrm{Z}_{sc}/\mathrm{Z}_{oc}|<1.$

Now

$$\frac{\mathrm{Z}_{sc}}{\mathrm{Z}_{oc}}=-\omega^2\mathrm{L}_m\mathrm{C}_m\left\{1-\frac{\mathrm{R}_m\mathrm{G}_m}{\omega^2\mathrm{L}_m\mathrm{C}_m}-j\left(\frac{\mathrm{R}_m}{\omega\mathrm{L}_m}+\frac{\mathrm{G}_m}{\omega\mathrm{C}_m}\right)\right\}, \quad . \quad . \quad (10)$$

and this can be written in the form

$$\frac{\mathrm{Z}_{sc}}{\mathrm{Z}_{oc}}=-r\{\cos\theta-j\sin\theta\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

whence

$$\mathrm{R}=\frac{\mathrm{R}_m}{l}\left[\sum_{n=1}^{\infty}\frac{(-1)^{n-1}r^{n-1}\cos(n-1)\theta}{(2n-1)}-\frac{\omega\mathrm{L}_m}{\mathrm{R}_m}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}r^n\sin n\theta}{(2n+1)}\right]. \quad (12)$$

and

$$\mathrm{L}=\frac{\mathrm{L}_m}{l}\left[\sum_{n=1}^{\infty}\frac{(-1)^{n-1}r^{n-1}\cos(n-1)\theta}{(2n-1)}+\frac{\mathrm{R}_m}{\omega\mathrm{L}_m}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}r^n\sin n\theta}{(2n+1)}\right]. \quad (13)$$

The series contained in these expressions are convergent for the values of "r" under consideration, and it is a simple matter to assign limits to the errors introduced by neglecting high-order terms. Upon substituting for "r" and "θ" and making suitable approximations, the following approximate expressions for R and L are obtained

$$\mathrm{R}=\frac{\mathrm{R}_m}{l}\left[1-\frac{\omega^2\mathrm{L}_m\mathrm{C}_m}{3}\left(2+\frac{\mathrm{L}_m\mathrm{G}_m}{\mathrm{C}_m\mathrm{R}_m}\right)+\frac{(\omega^2\mathrm{L}_m\mathrm{C}_m)^2}{5}\left(3+\frac{2\mathrm{L}_m\mathrm{G}_m}{\mathrm{C}_m\mathrm{R}_m}\right)\right.$$
$$\left.-\frac{(\omega^2\mathrm{L}_m\mathrm{C}_m)^3}{7}\left(4+3\frac{\mathrm{L}_m\mathrm{G}_m}{\mathrm{C}_m\mathrm{R}_m}\right)\right], \quad (14)$$

and

$$\mathrm{L}=\frac{\mathrm{L}_m}{l}\left[1-\frac{\omega^2\mathrm{L}_m\mathrm{C}_m}{3}\left(1-\frac{2\mathrm{R}_m\mathrm{G}_m}{\omega^2\mathrm{L}_m\mathrm{C}_m}-\frac{\mathrm{R}^2{}_m}{\omega^2{}_m\mathrm{L}^2{}_m}\right)\right.$$
$$\left.+\frac{(\omega^2\mathrm{L}_m\mathrm{C}_m)^2}{5}-\frac{(\omega^2\mathrm{L}_m\mathrm{C}_m)^3}{7}\right], \quad . \quad . \quad . \quad . \quad . \quad (15)$$

providing $\mathrm{L}_m\mathrm{G}_m/\mathrm{C}_m\mathrm{R}_m \not> 1$, a condition which is almost always satisfied in practice. Both equation (14) and equation (15) are accurate to within 0·1 per cent. when $\omega^2\mathrm{L}_m\mathrm{C}_m$ is not greater than 0·2, and when $\mathrm{R}_m/\omega\mathrm{L}_m$ and $\mathrm{G}_m/\omega\mathrm{C}_m$ are not greater than 0·1. It is interesting to note that, in these circumstances,

$$\mathrm{L}=\frac{\mathrm{L}_m}{l}\left\{1-\frac{\omega^2\mathrm{L}_m\mathrm{C}_m}{3}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

to within 1 per cent.

Leakance and Capacitance.

From equations (2), (3) and (4) the leakance and capacitance are seen to be given by

$$\mathrm{G}+j\omega\mathrm{C}=\frac{1}{\mathrm{Z}_{oc}l}\sum_{n=1}^{\infty}\frac{1}{(2n-1)}\cdot\left(\frac{\mathrm{Z}_{sc}}{\mathrm{Z}_{oc}}\right)^{n-1}. \quad . \quad . \quad . \quad . \quad . \quad (17)$$

Substituting in this equation from equation (11) gives

$$G=\frac{G_m}{l}\left[\sum_{n=1}^{\infty}\frac{(-1)^{n-1}r^{n-1}\cos(n-1)\theta}{(2n-1)}-\frac{\omega C_m}{G_m}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}r^n\sin n\theta}{(2n+1)}\right]. \tag{18}$$

and

$$C=\frac{C_m}{l}\left[\sum_{n=1}^{\infty}\frac{(-1)^{n-1}r^{n-1}\cos(n-1)\theta}{(2n-1)}+\frac{G_m}{\omega C_m}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}r^n\sin n\theta}{(2n+1)}\right]. \tag{19}$$

From these equations, when suitable approximations are made, are obtained

$$\begin{aligned}G=\frac{G_m}{l}\Bigg[1&-\frac{\omega^2L_mC_m}{3}\left(2+\frac{C_mR_m}{L_mG_m}\right)+\frac{(\omega^2L_mC_m)^2}{5}\left(3+\frac{2C_mR_m}{L_mG_m}\right)\\&-\frac{(\omega^2L_mC_m)^3}{7}\left(4+\frac{3C_mR_m}{L_mG_m}\right)+\frac{(\omega^2L_mC_m)^5}{9}\left(5+\frac{4C_mR_m}{L_mG_m}\right)\Bigg].\end{aligned} \tag{20}$$

$$\begin{aligned}C=\frac{C_m}{l}\Bigg[1&-\frac{\omega^2L_mC_m}{3}\left(1-\frac{2R_mG_m}{\omega^2L_mC_m}-\frac{G_m{}^2}{\omega^2L_m{}^2}\right)\\&+\frac{(\omega^2L_mC_m)^2}{5}-\frac{(\omega^2L_mC_m)^3}{7}\Bigg]. \quad . \quad . \quad . \quad .\end{aligned} \tag{21}$$

For all values of $\omega^2L_mC_m$ not greater than 0·2 and $R_m/\omega L_m$ and $G_m/\omega C_m$ not greater than 0·1, equation (20) is accurate to within 0·2 per cent., providing the ratio of $R_m/\omega L_m$ to $G_m/\omega C_m$ is not greater than 10, and equation (21) is accurate to within 0·1 per cent. Here again it is interesting to note that, in these circumstances,

$$C=\frac{C_m}{l}\left\{1-\frac{\omega^2L_mC_m}{3}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \tag{22}$$

to within 1 per cent.

It will be observed that most of the terms of equations (14), (15), (20) and (21) are common to all, and that the work involved in evaluating these four expressions is only slightly greater than that involved in evaluating any one of them. In general, the results will not be required to 0·1 per cent. accuracy and some terms may be neglected with further reduction in the labour of computation.

Having obtained the primary constants from the open- and closed-circuit impedances, by means of the above approximate expressions, the attenuation and phase constants can be calculated from the following expressions, which can be shown to hold for transmission lines at radio frequencies :—

The attenuation constant $=\frac{R}{2}\sqrt{\frac{C}{L}}+\frac{G}{2}\sqrt{\frac{L}{C}}$ nepers per unit length.

The phase constant $=\omega\sqrt{LC}$ radians per unit length.

Providing $R/\omega L$ and $G/\omega C$ are not greater than 0·1, the first of these expressions is accurate to within 0·2 per cent. and the second to within 0·5 per cent.

LXXVI. *On the Dimensions of Physical Magnitudes.* (Second Paper.)

By Herbert Dingle, A.R.C.S., D.I.C., D.Sc.*

[Received September 7, 1942.]

1. Before replying to Professor Wilson's comments † on my former paper under this title ‡, I wish to take this opportunity of repairing an omission. I should have mentioned Professor P. W. Bridgman's admirable book, 'Dimensional Analysis,' which is written from a viewpoint not indeed identical with mine, but so close to it as to make reference to the book almost inevitable. I read Bridgman's book many years ago, but when my interest in the subject was recently re-aroused from quite other sources, Bridgman had become associated in my mind with more general aspects of scientific theory, and I temporarily forgot that he had written on dimensions. On re-reading his book I realize what labour I should have saved myself had I kept it in mind. How far I am unconsciously indebted to him for the views expressed in my paper I do not, of course, know, but from my knowledge of the history of those views I suspect that, in the main, we have arrived independently at similar points of view—a result, I would fain believe, of the essential correctness of our philosophy.

2. I was pleased to notice that my paper had called forth comment from so valued a critic as Professor Wilson, but my pleasure was turned to disappointment when I found that he had concentrated attention almost entirely on utterly trivial points. Indeed, in these days of national obligations and paper shortage, I feel some compunction in following up such details—a course which can be justified, if at all, only by the necessity of meeting his general charge of "inaccuracies and misconceptions" which might otherwise prejudice consideration of the things that matter (I shall make this part of the discussion as brief as the need for finality allows), and the hope that I might be able to divert his attention to the essential points. There are two things about Professor Wilson s attitude that puzzle me very much. First, since he admits that he does not understand Dr. Brown's theory, and only "believes" he is "beginning to understand what Professor Dingle is groping after," why does he spread a "net" for Dr. Brown and charge me roundly with "inaccuracies and misconceptions"? Such action should follow, not precede understanding. I should have thought that he would have asked questions, not levelled accusations, and after I

---

* Communicated by the Author.

† Phil. Mag. xxxiii. p. 631 (1942).

‡ Phil. Mag. xxxiii. p. 321 (1942).

have answered his specific charges I shall accordingly interpret his paper as a somewhat unorthodox request for enlightenment and try to supply it. The second point is the remarkable disadvantage under which he places me by the assertion that I am " just the one natural philosopher who cannot afford to use language which is lacking in precision or meaning." Why am I unique in this respect, and where did he acquire such a stock of imprecision and meaninglessness that he can lavish it on me who alone can make no use of it ? This extra degree of liberty which he claims makes it very difficult for me to understand him ; indeed, I cannot be sure that he has not exercised it in the passage I have quoted. I am reminded of Meredith's critic who complained of a " tendency to polysyllabic phraseology " in the author he was reviewing.

3. My first " misconception " is that I have regarded the 2π in the formula

$$t=2\pi\sqrt{\frac{l}{g}}$$

for a simple pendulum as having something to do with angles. " Ever since I first felt that I had understood the credentials of this formula " writes Wilson, " I appear to have laboured under the misapprehension that the 2π in it has nothing whatever to do with angles." The time of swing of a simple pendulum is a function of the angular displacement, θ, of the bob. Dimensional analysis will yield the formula

$$t=f(\theta)\sqrt{\frac{l}{g}},$$

and a more particular analysis will determine $f(\theta)$ more precisely. Now if, first, we restrict ourselves to small values of θ, and, secondly, do not mind a slight inaccuracy, we find that $f(\theta)$ takes the particular value 2π. Has Professor Wilson always believed that when we give $f(\theta)$ a particular value, that value ceases to be a value of $f(\theta)$, or has he believed that $f(\theta)$ has nothing whatever to do with θ ? According to his statement he must have believed one or the other. I find it very difficult to credit this, but his statement is explicit, and so (unless, indeed, he is using his right to be imprecise or meaningless) I must accept it, and express my satisfaction at having had the honour of removing the misapprehension.

I should add that a reference to my former paper will show that the remark in question is robbed of its meaning by being taken from its context. In its proper place its significance will, I think, be understood, and it will be recognized as a quite incidental " aside." If Professor Wilson's confidence in my fundamental ideas, definitions and postulates is severely shaken by it, then by all means let it be eliminated. Its importance is negligible compared with that of the ideas, and I gladly remove it for the double gain of avoiding further quibbling and restoring Wilson's confidence to equilibrium.

4. The next point concerns precision of statement, and I am taken to task for looseness in my discussion of the definition of temperature. On

the general question, let me say at once that I acknowledge the obligation to be precise where precision is necessary for clear understanding, but I acknowledge no obligation to be precise for precision's sake. If I convey my meaning I care little for the foolish consistency that is the hobgoblin of little minds. For example, when, as here, I wish to contrast the definition of temperature as a corrected thermometer reading (*i. e.*, a fundamental magnitude) with the definition of temperature in terms of molecular energy (*i. e.*, a derived magnitude), the essential point is that the dimensional symbol is in the first case $[\theta]$ and in the second $\left[\frac{\mathrm{E}}{\mathrm{M}}\right]$, or $[\mathrm{L}^2\,\mathrm{T}^{-2}]$. So long as I can make that clear, I have no interest in giving every detail of the definitions in question.

Wilson, however, thinks that instead of indicating temperature according to the second definition as " proportional to the kinetic energy per unit mass," I should have made it " proportional to the average kinetic energy of translation per gram molecule." Well, if we are to be precise, it is worth pointing out that a quantity inversely proportional to a mass in grams is also inversely proportional to a mass in gram molecules. True, the constant of proportionality may vary, but that is quite immaterial for my purpose, and I have not said that it will not. However, to remove Professor Wilson's " difficulties," I would be prepared to accept the amendment were it not for the other part of the definition. Here I am required to substitute " average kinetic energy of translation " for " kinetic energy." Now precision would require " kinetic energy of translation reckoned with respect to a frame of reference in which the body as a whole is at rest." Wilson is perturbed at my omitting reference to translation, which introduces a small correcting factor at the most, but himself overlooks the necessity of stating the frame of reference, the omission of which leaves the temperature infinitely vague. The net result then, is that I, striving only to be understood, have succeeded, for Wilson, in spite of his " difficulties," is so confident of my meaning that he ventures to correct my expression of it ; whereas he, striving to be precise, has succeeded only in being as indefinite as it is possible to be. I will therefore, after all, not amend my original statement ; I will accept for this occasion the pragmatic philosophy which Professor Wilson recommends, and retain the statement which has worked.

5. Wilson touches a more important point when he disputes my assertion that the temperature in interstellar space according to the first definition is very different from that according to the second. He says " the reading ' 3° absolute,' has no meaning, so far as temperature is concerned," 3° absolute being the temperature which a hydrogen gas thermometer would be expected to record in the region in question. This is a surprising statement, notwithstanding that the estimate of 3° is not mine particularly, but that of physicists generally, who, according to Professor Wilson, can afford to be meaningless. It is surprising because the reason which Wilson gives for the meaninglessness of the 3°

absolute is that it does not refer to equilibrium conditions. But it does. You place the thermometer—an ordinary hydrogen thermometer with blackened bulb and, of course, a substance which remains liquid at the prevailing temperature substituted for mercury—in interstellar space, wait for its reading to become steady, apply the orthodox corrections, and there is your temperature—about 3° absolute.

Is that not a legitimate way of measuring temperature ? Let me quote Maxwell, the handiest authority I have whom I have reason to believe would be acceptable to Professor Wilson. He says *—" If we bring the thermometer into intimate contact with another substance, as for instance if we plunge it into liquid for a sufficient time, we find that the reading of the thermometer becomes higher or lower according as the liquid is hotter or colder than the thermometer, and that if we leave the thermometer in contact with the substance for a sufficient time the reading becomes stationary. Let us call this ultimate reading 'the temperature of the substance.' We shall find as we go on that we have a right to do so." In view of this I see no reason to modify the passage criticized.

6. Professor Wilson's comments on my section on electrical and magnetic magnitudes I find completely unintelligible. He says that his equations (13), giving dimensional formulæ for pole-strength (m) and permeability (μ), are implied by my definitions, but they certainly are not. I am not interested in pole-strength in this connexion, but I showed in my paper that the dimensions of μ, according to my scheme, were $[M^{-1}T^{3}]$. Wilson has made no attempt to find an error in my reasoning, which indeed is extremely simple. He has simply ignored my result and stated bluntly that the dimensions which I must give to μ are $[ML^{-1}T^{-2}]$. Until he shows why he thinks I must do so, I cannot help him out of his difficulty.

While on this point I might remark that the word " permeability " is used ambiguously in physics for both the dimensional constant appearing in my equation (49) and the (dimensionless) ratio of the fields created in a solenoid containing successively a material medium and a vacuum. My formula, of course, applied to the first meaning of the word. The corresponding ambiguity in electricity is somewhat relieved by the possibility of calling the dimensional quantity the " dielectric constant " and the other the " specific inductive capacity," and it would, I think, be desirable to make this distinction in nomenclature. I briefly discussed the relation between the two electrical quantities on the last page of my former paper, and a similar comment is applicable to the magnetic quantities.

7. And now, having, I hope, emancipated the discussion from trivialities, I would like to reciprocate the hope which Professor Wilson expresses for some measure of agreement between him and me, and I will therefore discuss one or two points, on which, from my reading of his

---

* 'Theory of Heat,' Second Edition, p. 5.

papers, I judge that we take different views. The first relates to what we mean by a physical magnitude. Take a length, for example. I regard what we call " the length of a body " as *the result of an operation* in which the body is compared in a prescribed way with a chosen " standard " body ; I do not regard it as an intrinsic property of the body, which is " discovered " by means of the standard. Wilson seems to take a different view, for in an earlier paper * he calls a length a " ratio of numbers obtained by similar measurements," implying that the body to be measured has an intrinsic length, that the standard also has an intrinsic length, and that " the length of the body " is the ratio of one of these lengths to the other. Such a view leads one to ask, " What is this length that both bodies possess ? " and there is no answer ; " length " remains a mystical conception. On my view there is no meaning in speaking of the length of the standard, because you cannot perform the operation, for the result of which the word " length " stands, with respect to it.

The dimensional symbol for length, namely [L], is then, on my view, a shorthand sign for the specified operation. That is clear and simple. On Wilson's view I am not sure what [L] means. Does it stand for the ratio of the lengths of the body and the standard ? That would seem to follow, but then how can one represent the length of the body and the length of the standard separately ? If each of these is also to be [L], then the ratio can scarcely be [L] too. With such basic conceptions it is no wonder that the theory of dimensions takes on the mystical aspect in which Wilson sees it.

In a similar way we perform other operations, which we call the measurement of time, of mass, . . ., and represent them by symbols such as [T], [M], . . . We then form certain algebraical combinations of the results of our operations, and assign appropriate dimensional symbols to them. Thus, the combination $[ML^{-3}]$ is very useful to us : we call it *density*, and the dimensional symbol indicates just what operations we perform in measuring density, and how their results are combined. (It might be desirable to say that I am aware that indirect equivalents of these operations are often substituted for them for practical reasons.) The symbol is thus always a shorthand sign, representing our definition (not a mystical aura) of the quantity for which it stands.

All these operations are, of course, chosen by us out of the infinite number of curious things we might do, and those survive which prove helpful to us in ordering our experience. Unless we are working in a trance we should never be at a loss for the symbol for any magnitude, for, having chosen the operation which yields the magnitude, we have only to write the process down in shorthand form. To ask, for instance, what are the dimensions of viscosity is to ask what the word " viscosity " stands for. It did not descend upon us from heaven ; we formed it to stand for something we chose to concern ourselves with, and the dimensional symbol must simply indicate how we formed it. To say that the dimensions of viscosity are inaccessible is simply to say that we do not know what we are doing. Wilson seems to imply, in the paper referred to above, that the dimensions of κ and μ are inaccessible because every electromagnetic relationship is equivalent to a single equation for two unknowns. This I regard as a complete misapprehension. The same basic error comes to light again in Wilson's latest paper, when he remarks that from my definition of electric charge,

---

* Phil. Mag. xxxiii. p. 29 (1942).

$$q=\mathrm{N}f,$$

I "infer, rather temerariously,"

$$[\mathrm{Q}]=[\mathrm{F}].$$

This is no inference and there is nothing temerarious about it ; it merely expresses in symbols the definition chosen. If there is any temerity (I acknowledge none) it is in the choice of the definition of charge ; there cannot be any in merely expressing the choice in symbols instead of words.

The foregoing remarks, however, do not require that to every physical quantity there is a single, unique dimensional symbol. The symbol uniquely represents the definition of the quantity for whose magnitude it stands, but the definition is not unique ; we have a choice. For example, we can define mass and length by single operations, and represent these operations by [M] and [L]. We can then define volume as the combination represented by $[\mathrm{L}^3]$, and density as the combination represented by $[\mathrm{ML}^{-3}]$. This, in fact, is the procedure which is commonest in modern physics. But we can, if we wish, choose an independent operation for defining density (*e. g.*, the use of a hydrometer). We should then represent density by the symbol [D]. We might then give up our former definition of mass, and re-define it as the product of density and volume, whereupon it would acquire the symbol $[\mathrm{DL}^3]$. This is the procedure implied in Newton's definition of mass as "the measure of the quantity of matter arising from its density and bulk conjointly."

There is nothing whatever in nature to decide which scheme we shall choose. To ask whether [M] or $[\mathrm{DL}^3]$ is the "correct" symbol for mass is to reveal a complete misapprehension of the meaning of dimensional symbols. There is no unique "correct" symbol for any magnitude, though there may, of course, be incorrect ones, their incorrectness arising from the fact that the symbols do not represent the definitions, or from a choice of definitions which are not self-consistent. Wilson makes several references to the "correct" dimensions of magnitudes, which seem to imply that his ideas on this subject are not altogether clear. I should like to assure him that my suggested scheme of electrical and magnetic magnitudes was not at all intended to be the one and only correct one. All I claim for it is that it is *a* correct one (or, so far as I carried it, the basis of a correct one), and in that respect it is superior to the one usually given in text-books, according to which we cannot assign dimensions to κ and μ separately, and thereby acknowledge that we do not remember what we have invented. If Wilson wishes to challenge my scheme he must either find some inconsistency in it, or else show that the definitions I have proposed give values of the corresponding quantities incompatible with those we actually use. To prove that my equation

$$[\mu\kappa]=[\mathrm{L}^{-1}\,\mathrm{T}],$$

which " shocks " him, is different from the " correct " one

$$[\mu^{-\frac{1}{2}}\kappa^{-\frac{1}{2}}]=[\mathrm{LT}^{-1}]$$

is idle ; I knew that before.

Circumstances may arise, however, in which it becomes possible to speak of the " correct " dimensions—namely, when the definition of the quantity in question has been decided by general agreement. The correctness then arises from convention, not from necessity. The desirability of such agreement was emphasized in Section III. of my former paper, and arises from the fact that alternative schemes which, with limited knowledge, appear equivalent, may cease to be so when wider experience is obtained, and confusion and error may thus be caused by an unconscious change from one to another. I gave two examples to illustrate this—the measurements of time and of temperature. These examples are not the only ones possible, though, so far as I know, they are the most serious. " What a pity," writes Professor Wilson, " he did not think of cases like the following :

$$\text{Heat}=\mathrm{N}\times\text{Work.''}$$

I assure him that I did, but, as I remarked, thermodynamical magnitudes illustrate nothing new in principle, so I saw no point in labouring what I thought already adequately exemplified. Here, however, if it helps Professor Wilson, is the example which seems to appeal to him most readily.

Heat was originally defined in terms of change of temperature, and the unit of heat was the amount which would raise 1 gm. of water through 1° C. (I am again going to be imprecise, and omit details as to which degree, conditions of pressure, etc.) The dimensions of heat were therefore given by

$$[\mathrm{H}]=[\mathrm{M}\theta],$$

for the magnitudes of mass and temperature were the only ones, a change in the measurement of which would affect the measurement of a given quantity of heat. It was then discovered that whenever a transformation of heat, as so defined, into work, or *vice versa*, occurred, the amount of the one which appeared was always proportional to the amount of the other which disappeared. We thus had the empirical relation

$$\mathrm{W}=\mathrm{JH}$$

or

$$[\mathrm{ML}^2\mathrm{T}^{-2}]=[\mathrm{JM}\theta],$$

giving J as a *dimensional* constant having the dimensions $L^2T^{-2}\theta^{-1}$. But this soon led to the tacit introduction of a *new* measure of heat, in terms of the work into which it could be transformed. We then had precisely Wilson's equation, with N=1. This, and no other definition, is necessary to save from dimensional absurdity the common expression of the First Law of Thermodynamics—

$$dE=dW-dH.$$

The old definition would, of course, require

$$dE=dW-JdH,$$

with J having the dimensions already given. We thus have two distinct definitions of heat measurement at large in physics, with all the potential dangers inherent in such a situation. There is, we believe, no actual danger, because we have faith in the universality of Joule's law, but in principle the position is just as unsatisfactory as that involved in my other examples.

8. There is nothing, of course, in the preceding paragraph that is not already contained in the postulates I set out earlier. My excuse for the repetition is that I hope that I have thereby brought my view of the theory of dimensions into direct relation with those predispositions of Professor Wilson's, which I judge to conflict with it.

LXXVII. *Erratum to the Paper "The Formula of the Selenium Barrier-Layer Photo-cell."* Phil. Mag. p. 226 (March 1942).

To the Editors of the Philosophical Magazine.

Gentlemen,

In the third sentence of the above paper I state that no one seems to have previously investigated the formula which gives the current in the above cell as a function of the illumination and external resistance. This is inaccurate; the formula has been investigated for eight cells in the G. E. C. Laboratories at Wembley by J. R. Atkinson, N. R. Campbell, E. H. Palmer, and G. T. Winch (Proc. Phys. Soc. l. p. 934, 1938).

Yours faithfully,
R. A. Houstoun.

August 10, 1942.

LXXVIII. *Notices respecting New Books.*

Introduction to Atomic Physics. By S. Tolansky. [Pp. xii+343+5 plates.] (Longmans, Green & Co., 1942. Price 15*s.*)

This book, in the words of the author in the preface, " . . . . aims at introducing in an elementary manner a fairly comprehensive survey of modern atomic physics to students who have already passed through an intermediate course." Based on a course of lectures delivered to Honours Physics students in their penultimate year and final year General Honours Science students, it contains a complete, qualitative exposition of the major branches of modern physics. In connexion with the building-up of theory, emphasis is placed on experimental evidence and resort is had to mathematical description only when clarity of exposition makes it desirable.

Not all subjects which could properly be classed as branches of atomic physics are included, but nevertheless Dr. Tolansky has covered a very wide range. He deals thoroughly and competently with X-rays and crystal structure, quantum theory, atomic theory and spectra, and includes a brief introduction to wave mechanics. The thermionic effect, photo-electricity and the conduction of electricity through gases are admirably surveyed. A third of the book is devoted to an excellent account of nuclear physics—α-, β- and γ-rays, the neutron, positron and meson, and nuclear structure. Finally, there is a satisfying chapter on the special theory of relativity.

The book is very well produced and there are few typographical errors which are not covered by a list of errata. It is rather unfortunate that the general excellence of the book should be marred by an incomprehensible departure from standard mathematical notation in the constant use of the difference symbol δ for the partial derivative symbol ∂. It is very disturbing to find Poisson's equation given as $\delta^2V/\delta x^2=4\pi\rho$.

Dr. Tolansky is to be congratulated on having succeeded so admirably in attaining his aim. This volume essentially provides the student with a sound and comprehensive knowledge of the results and theories of modern atomic physics without burdening him with mathematical detail. The chapters on nuclear physics are specially to be commended. The book should be warmly welcomed by both students and teachers. S. R. T.

Introduction to Physical Chemistry. By ALEXANDER FINDLAY. Second Edition, revised and enlarged. [Pp. vii+582.] (Longmans, 1942. Price 15*s.*)

THIS well-known and widely used introductory text-book has been increased in size by 90 pages. It has been thoroughly revised, and in parts rewritten in such a way that the new material has been skilfully grafted on to the old. The main plan is the same as in the first edition, but the chapters on Colloids and Adsorption have been moved forward to an earlier stage in the treatment, and a short new chapter on Molecular Dimensions and Configurations has been added. The chapter on the Constitution of Matter has been extended and considerably improved, and that on the Solid State is now double its former length. The new material includes the following topics :—transmutation, electronic arrangement and the periodic table, the methods for the determination of exact molecular weights by limiting densities and by use of the microbalance, the maximum bubble pressure method for the determination of surface tension, absorption spectra of liquids, molecular association, atomic and molecular heat, the elements of crystallography, the determination of the space lattice and the applications of X-ray analysis, interatomic distances and the relative direction of bonds, the dew-point method for the determination of molecular weight, the osmotic coefficient, the antimony electrode, the glass electrode, oxidation-reduction indicators, the determination of the dissociation constants of acids by potential measurements, the mechanism of reactions in the gaseous and the liquid phases, energy of activation, contact catalysis, chain reactions, the temperature coefficient of photo-reactions, the ultracentrifuge, adsorption indicators and unimolecular films. Many minor alterations and additions have also been made. The thermodynamic treatment has been tightened up and improved ; the recommended symbols for the thermodynamic functions are used and there is now a very brief mention of entropy. More worked numerical examples have been included in the text—a feature which might be extended with advantage in future editions ; the number of problems at the end of the book has been doubled and a short table of physical constants added.

As far as is possible the historical method of treatment has been adopted, as this approach is better suited to the needs of the elementary student ; the result is that the outlook of the book is that of classical physical chemistry brought up to date. New methods and modern developments are, however, best appreciated by those who have received a sufficient grounding in the old, and such a sound introduction is provided by this new edition. There are references to monographs and articles for better students whose appetite has been whetted for more. It is a pity that the price has had to be increased, but it is still no higher than that of most text-books of similar scope. J. W. B.

A History of Science and its Relations with Philosophy and Religion. By Sir William Cecil Dampier. Third Edition. [Pp. xxiii+574.] (Cambridge University Press, 1942. Price 25*s*.)

In his 'History of Science,' first published in 1929, Dampier admirably fulfilled the need for "another attempt," as he put it in his preface, "to tell the general story of science on Whewell's lines, to present, not a detailed study of one period or subject, but a complete outline of the development of scientific thought." A second edition, required within a few months of the first, has been out of print for some time. For the new edition no drastic changes have been made in the original text, but a number of minor modifications have been made in the course of the revision, and some references have been added. There is a new 50-page chapter carrying the story of the development of science over the last ten years.

The new chapter, to which there is a temptation to turn first, deals with a wide range of topics, among them alloys and solutions, the nuclear atom, geophysics, enzymes, vitamins, viruses, genetics, and Eddington's philosophy of the physical sciences. It is a chapter in which a reader may catch exciting glimpses of fields other than those with which he is familiar, and he will be grateful to the author for writing it. Written "in the intervals of more urgent work" it is, however, inevitably lacking in cohesion, and would, perhaps, have been better as an appendix. As another chapter it comes oddly after the artistic rounding off of the pattern of the book with the survey of scientific philosophy and its outlook.

It is hardly necessary to call attention to the many excellences of the broad survey of the development of the sciences from the earliest times to the present day in a book already well known. Of that survey, reflecting the author's wide knowledge of, and enthusiasm for, both the physical and biological sciences, there could surely be nothing but unstinted admiration. About the treatment of the wider relations of science, however, there may well be divergence of opinion. There is a balanced presentation of what may be called the classical philosophical positions, but for more recent times the account becomes somewhat one-sided. The author may not rate logical positivism or dialectical materialism very highly, but no general account of the recent relations between science and philosophy which does not refer to them can be considered entirely satisfactory. More important than this, the author seems to ignore almost completely the interdependence, in the sense in which it has been stressed in recent years, of science and social and economic conditions, the appreciation of which, it has been suggested, "has transformed the history of science from a minor to a major subject."

"Those working at specific scientific problems," Dampier says, "have perhaps no need of history, but those who try to understand the deeper meaning of science itself, and its connexion with other subjects of human thought and activity, must know something of the story of its development." Criticism of a book for what it lacks is not inconsistent with praise for what it includes. What this book includes essentially is an admirably written history of science and related thought, such as should be read by every student of science and every practising scientist, each adding to what it gives such commentary and interpretation as his own experience makes possible. E. C. S.

Reports on Progress in Physics.—Vol. VIII (1941). [Pp. iv+368.] General Editor: W. B. Mann. (London: The Physical Society, 1942. Price 25*s.*)

It is a pleasure at the present time to find that a volume reporting progress indicates in a manner worthy of its predecessors in this series that movement in a forward direction is still appreciable. The policy of surveys of selected topics covering a number of years' work is again justified over that of the year-book.

This volume opens with a wise and very witty article by M. Born on the teaching of theoretical physics in Universities. M. Blackman discusses recent developments in the theory of specific heats of solids, particularly those arising out of the new results regarding the vibrational spectra. A reconsideration of photometric ideas and standards has long been overdue, so that H. Buckley's excellent article is welcome. Theoretical and practical spectroscopy take up a considerable proportion of the book in the articles of A. G. Gaydon on "Applications of Spectroscopy of Combustion," A. Hunter on "Instrumental Technique in Astrophysics," Lotte Kellner on "Dispersion in the far Infra-Red," R. S. Mulliken and C. A. Rieke on "Molecular Electronic Spectra, Dispersion and Polarization," and G. R. Harrison on "The M.I.T. Wave-length Project." English physicists will, no doubt, gasp with admiration and envy at the thoroughness of the M.I.T. in reducing the measurement and classification of spectra to a matter of pressing the button of a recording and sorting machine.

This volume illustrates too the peculiar value of this series in calling the attention of physicists to advances in sometimes forgotten corners, as in the articles by J. W. Beams on "High-speed Centrifuging," by R. Schnurmann on "Friction between Solid Bodies," and the critical survey by R. W. Lunt, A. von Engel and J. M. Meed of present ideas on the mechanism of discharges in gases. Theoretical physicists will be interested in the papers by R. T. Birge on "The General Physical Constants," and by E. P. Wigner and E. Feenberg on "Symmetry Properties of Nuclear Levels," while the highly technical account of recent television developments will no doubt be useful to those who are specially concerned. J. E.

Spectroscopy and Combustion Theory. By A. G. Gaydon, M.Sc., Ph.D. With a foreword by A. C. Egerton, F.R.S. [Pp. vii+191.] (London: Chapman and Hall, 1942. Price 17*s.* 6*d.*)

The bulk of this monograph is concerned with recent work on the spectra of flames and explosions and its interpretation. The survey of experimental results is skilfully balanced; the critical discussion of their significance clear and concise. In a field where the absence of decisive evidence makes vague speculation almost irresistible, Dr. Gaydon maintains an admirable restraint. Throughout, his evident mastery of the technique of the theory and practice of molecular spectroscopy and his power to communicate his own enthusiasm makes the reading most pleasurable. The chapters on the life-times of activated molecules, the determination of flame temperatures, the lag in the equipartition of energy liberated by combustion processes and the calculation of heats of dissociation, will be found most valuable to those whose interest is of a more general kind. The evidence for Dr. Gaydon's own theory of the after burning of carbon monoxide is marshalled effectively. The student and workers in other fields will find this book readable and stimulating, especially as their needs are considered in introductory chapters which explain adequately in non-mathematical language sufficient of the theory of molecular spectra for a complete understanding of the rest. Research workers will need no intimation that this is a volume not to be missed, particularly as its intrinsic value is increased by a list of two hundred and fifty-eight references given conveniently at the end of the book. J. E.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

LXXIX. *Elastic Waves and Vibrations of Thin Rods.*

By JOHN PRESCOTT, M.A., D.Sc.*

[Received March 30, 1942.]

THE general equations of motion of a homogeneous elastic solid lead to the conclusion that, in an infinite solid at least, two different kinds of wave can be transmitted through the body at the same time, each with its own velocity. Anything that can happen in an infinite body can also happen in a finite body, the only difference being the effect of the boundaries of the finite body in reflecting the waves back into the body, and, until a wave reaches a boundary, it is, in effect, travelling in an infinite solid. The two types of waves are known respectively as *dilatational* (the sort that exist in a gas in sound waves) and *distortional* or *rotational*. It is infinitely improbable that one of these types exists without the presence of the other in any actual disturbance in an elastic body, but there are certain ideal problems in which they can exist separately. The distortional type exists in all its simplicity in the simplest purely torsional oscillations of a cylindrical (circular) rod ; and if an explosion occurred in a spherical hole inside a homogeneous body then purely dilatational waves could be set up as a result of this.

It is proved that plane and spherical waves, of either type, travel with constant velocities and, in the case of plane waves, the wave changes neither its form nor its intensity as it travels. That is, the same sequence of disturbances in the same order and at the same intervals of time occur at successive points along the normal to the plane of the wave. In the case of spherical waves the disturbance travels, of course, along radii and again the wave form is preserved, but with diminished intensity as the wave moves outwards from the centre. This means that, although the curve representing the wave has the same shape at all times, the ordinates of the curve decrease as the radius r increases.

For future reference we give here the meanings of some of the symbols we shall use :—

E=Young's modulus.
n=rigidity modulus.
σ=Poisson's ratio.
$m=n/(1-2\sigma)$.
ρ=density of material.
V_D=velocity of dilatational waves.

---

* Communicated by the Author.

V_n=velocity of rotation waves.

$V_D{}^2=(m+n)/\rho\,;\ \ V_n{}^2=n/\rho\,;\ \ V_E{}^2=E/\rho.$

λ=wave-length of sine waves.

M=bending moment in rod.

F=shear force on normal section of rod.

Since there are only two independent elastic constants in a homogeneous solid there exist relations between more than two. Thus

$$E/n=2(1+\sigma),\quad E/(m+n)=1-\frac{2\sigma^2}{1-\sigma}.$$

The following systems of co-ordinates and corresponding displacements and stresses will be used :—

Co-ordinates	x	y	z	x	r	θ
Displacements	u	v	w	u	s	η
Tensile stresses.....	P_x	P_y	P_z	P_x	P_r	P_θ
Shear stresses	S_x	S_y	S_z	S_x	S_r	S_θ

$$\Delta=\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}.$$

The angle θ will be measured from the y axes, so that

$$y=r\cos\theta,\quad z=r\sin\theta.$$

The axis of x will always be taken along the undisturbed axis of a thin rod, and, in transverse motion, the central line of the rod (the line of centroids) will remain in the xz plane. In the usual theory of thin rods we should say that the transverse motion of the axis of the rod is in the direction of the z-axis.

Dilatational motion is given by

$$u=\frac{\partial\phi}{\partial x},\quad v=\frac{\partial\phi}{\partial y},\quad w=\frac{\partial\phi}{\partial z}, \qquad (1)$$

where ϕ is a function of x, y, z, t, satisfying the equation

$$V_D{}^2\nabla^2\phi=\frac{\partial^2\phi}{\partial t^2}. \qquad (2)$$

A plane wave, travelling in the positive x direction, is completely expressed by

$$\phi=f(x-V_Dt), \qquad (3)$$

where f denotes any function whatever.

For this wave

$$u=f'(x-V_Dt),\quad v=0,\quad w=0. \qquad (4)$$

Thus ϕ expresses a disturbance which remains constant when $(x-V_Dt)$ remains constant ; that is, a disturbance which remains the same but travels in the x direction with a velocity V_D.

It is easy to show that, if ϕ is a function of R and t only, where

$$R^2=x^2+y^2+z^2,$$

then the spherical waves are expressed by

$$\phi = \frac{1}{\mathrm{R}} \mathrm{F}(\mathrm{R} - \mathrm{V}_{\mathrm{D}} t). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The radial displacement u_{R} corresponding to this is

$$u_{\mathrm{R}} = \frac{\partial \phi}{\partial \mathrm{R}} = \frac{1}{\mathrm{R}} \mathrm{F}'(\mathrm{R} - \mathrm{V}_{\mathrm{D}} t) - \frac{1}{\mathrm{R}^2} \mathrm{F}(\mathrm{R} - \mathrm{V}_{\mathrm{D}} t),$$

and there is no other displacement.

Thus for plane waves ϕ remains the same as the wave travels forward, whereas it is $\mathrm{R}\phi$ that remains the same for spherical waves. It was, of course, only to be expected from energy principles that the intensity of a spherical wave would decrease as R increases, and this is the significance of the factor R^{-1} in the expression for ϕ.

Distortional or rotational waves are obtained from a similar set of equations. The displacements due to distortion satisfy the equations

$$\left. \begin{aligned} \mathrm{V}_n^2 \nabla^2 u &= \frac{\partial^2 u}{\partial t^2}, \\ \mathrm{V}_n^2 \nabla^2 v &= \frac{\partial^2 v}{\partial t^2}, \\ \mathrm{V}_n^2 \nabla^2 w &= \frac{\partial^2 w}{\partial t^2}, \end{aligned} \right\} ; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

together with the condition that the dilatation

$$\Delta \equiv \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

is zero. This last condition reduces the three equations (6) to two independent equations, and it is easy to represent u, v, w, by means of two functions.

The same remarks concerning plane and spherical waves apply to rotational motion as to dilatational, since the differential equations are the same. The difference is that the velocity is V_n instead of V_{D} ; and

$$\mathrm{V}_{\mathrm{D}}^2/\mathrm{V}_n^2 = \frac{m+n}{n} = \frac{2-2\sigma}{1-2\sigma} = \frac{7}{2} \text{ for } \sigma = 0{\cdot}3. \quad . \quad . \quad . \quad (8)$$

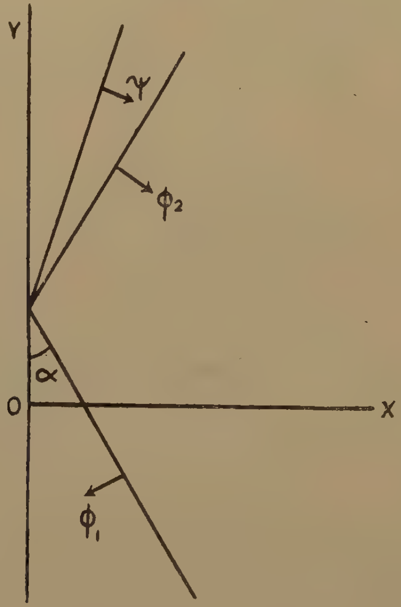
(In all our arithmetical work we shall take $\sigma = 0{\cdot}3$.)

What theory tells us then is that any kind of disturbance in an elastic body travels with a velocity V_{D} if it is of dilatational type and with velocity V_n if it is of rotational type, and the two types can co-exist and not interfere with each other. Further, each element of disturbance, or state of stress at any point, is transmitted by spherical waves. The plane wave is only the wave front of an infinity of spherical waves whose centres are on another plane parallel to the wave, which may be regarded as the plane of origin of the whole disturbance.

Poisson proved that the dilatational disturbance at any point P at any instant is merely the mean of the disturbances over the surface of a sphere of radius $V_D t$, with centre at P, at a time t seconds before the instant considered. This applies to an infinite solid and to any regular or irregular kind of disturbance. This shows that, as we remarked above, every element of disturbance spreads itself out in spherical waves moving with a velocity V_D in straight lines. The same is true for rotational disturbances with the substitution of V_n for V_D.

When an elastic disturbance is transmitted in a body which has free boundaries, a wave, on meeting a boundary, is reflected back into the solid in an altered form. A dilatational wave is reflected, as a rule, partly as a dilatational and partly as a rotational wave. The easiest case to consider is a plane wave striking a plane boundary obliquely.

Fig. 1.



It will be found that the boundary conditions cannot be satisfied without the use of reflected waves of both kinds, dilatational and rotational. For the sake of getting a clear picture of how reflexion takes place, it is worth while working out this simple problem in detail.

We assume that a dilatational wave

$$\phi=\phi_1(ct+\lambda x+\mu y) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

approaches the free boundary $x=0$ from the side where x is positive. Here we use c for V_D, and

$$\lambda=\cos\alpha, \quad \mu=\sin\alpha, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

α being the angle which the direction of motion of the wave makes with the negative-x direction (see fig. 1). (λ does not mean a wave-length here.)

The displacements due to ϕ_1 are

$$u_1=\frac{\partial\phi_1}{\partial x}=\lambda\phi_1'(ct+\lambda x+\mu y), \tag{11}$$

$$v_1=\frac{\partial\phi_1}{\partial y}=\mu\phi_1'(ct+\lambda x+\mu y), \tag{12}$$

and the stresses that come into the boundary conditions are

$$\begin{aligned} \mathrm{P}_x&=(m+n)\frac{\partial u_1}{\partial x}+(m-n)\frac{\partial v_1}{\partial y}, \\ &=a\frac{\partial u_1}{dx}+b\frac{\partial v_1}{\partial y} \text{ say,} \\ &=(a\lambda^2+b\mu^2)\phi_1''(ct+\lambda x+\mu y)\,; \end{aligned} \tag{13}$$

$$\begin{aligned} \mathrm{S}_z&=n\left(\frac{\partial u_1}{\partial y}+\frac{\partial v_1}{\partial x}\right), \\ &=2n\lambda\mu\phi_1''(ct+\lambda x+\mu y). \end{aligned} \tag{14}$$

At $x=0$ the stresses involve only the function $\phi''(ct+\mu y)$. It is easy to see, therefore, that the reflected dilatation wave necessary to balance the stresses of the incoming wave has the form

$$\phi=\phi_2(ct-\lambda x+\mu y), \tag{15}$$

which gives stresses at the free surface

$$\mathrm{P}_x=(a\lambda^2+b\mu^2)\phi_2''(ct+\mu y), \tag{16}$$

$$\mathrm{S}_z=-2n\lambda\mu\phi_2''(ct+\mu y). \tag{17}$$

On a mere inspection of these stresses it can be seen that the total stresses due to ϕ_1 and ϕ_2 cannot both be zero for any choice of ϕ_2. So we need a rotation wave as well.

Now the rotation equations (6) are clearly satisfied by

$$u=\frac{\partial\psi}{\partial y},\quad v=-\frac{\partial\psi}{\partial x},\quad w=0, \tag{18}$$

where ψ is a function satisfying

$$\mathrm{V}_n{}^2\nabla^2\psi=\frac{\partial^2\psi}{\partial t^2}. \tag{19}$$

A plane wave travelling in the positive x direction is represented by

$$\psi=\psi(\mathrm{V}_nt-\lambda_3x+\mu_3y). \tag{20}$$

The stresses corresponding to this function at $x=0$ are

$$\begin{aligned} \mathrm{P}_x&=a\frac{\partial^2\psi}{\partial x\,\partial y}-b\frac{\partial^2\psi}{\partial y\partial x}, \\ &=-\lambda_3\mu_3(a-b)\psi''(\mathrm{V}_nt+\mu_3y), \\ &=-2n\lambda_3\mu_3\psi''(\mathrm{V}_nt+\mu_3y)\,; \end{aligned} \tag{21}$$

$$\begin{aligned} \mathrm{S}_z&=n\left(\frac{\partial^2\psi}{\partial y^2}-\frac{\partial^2\psi}{\partial x^2}\right), \\ &=n(\mu_3{}^2-\lambda_3{}^2)\psi''(\mathrm{V}_nt+\mu_3y). \end{aligned} \tag{22}$$

If $f=V_n/V_D=V_n/c$, we can make the function ψ identical with the function ϕ, except for a constant factor, provided we put

$$\mu_3=f\mu. \qquad (23)$$

Then

$$\psi(V_nt+\mu_3y)=\psi\{f(ct+\mu y)\} \qquad (24)$$

and

$$\psi''(V_nt+\mu_3y)=\frac{1}{f^2}\frac{d^2\psi}{d^2(ct+\mu y)}. \qquad (25)$$

Now, collecting up our stresses and equating their sums to zero at the boundary $x=0$ we get, omitting the arguments $(ct+\mu y)$ for the ϕ functions and $f(ct+\mu y)$ for the ψ function,

$$(a\lambda^2+b\mu^2)(\phi_1''+\phi_2'')-2n\lambda_3\mu_3\psi''=0 \qquad (26)$$

and

$$2n\lambda\mu(\phi_1''-\phi_2'')+n(\mu_3^2-\lambda_3^2)\psi''=0. \qquad (27)$$

These can now be satisfied and give results of the forms

$$\phi_2''=B\phi_1'',$$
$$\psi''=C\phi_1'',$$

from which we find

$$\phi_2=B\phi_1, \qquad (28)$$
$$\psi=Cf^2\phi_1,$$
$$=Cf^2\phi_1(ct+\mu y). \qquad (29)$$

The complete ψ function is thus

$$\psi=Cf^2\phi_1\left(ct+\mu y-\frac{\lambda_3}{f}x\right),$$
$$=Cf^2\phi_1\left(\frac{V_nt+\mu_3y-\lambda_3x}{f}\right). \qquad (30)$$

Thus the original plane dilatation wave arriving at the free boundary gives rise to two plane reflected waves, a dilatation wave and a rotation wave. The reflected wave of the same type (dilatation) is reflected from the free surface at the angle of incidence; but the relation

$$\frac{\mu_3}{\mu}=f=\frac{V_n}{V_D}$$

shows that the rotation wave is subject to the same rule as refracted light waves going from one medium to another, in which the velocities of propagation are different.

If the original wave that arrived at the free boundary $x=0$ had been the rotation wave

$$\psi_1(V_nt+x\cos\alpha+y\sin\alpha),$$

it would be found that the reflected waves, rotational and dilatational, would have been

$$\psi_2(V_nt-x\cos\alpha+y\sin\alpha)$$

and

$$\phi(V_Dt-\lambda x+\mu y),$$

where

$$\lambda^2+\mu^2=1, \qquad (22a)$$

and

$$\mu=\frac{V_D}{V_n}\sin\alpha. \qquad (23a)$$

The relations between the functions are, at $x=0$,

$$2n \sin \alpha \cos \alpha(\psi_1''-\psi_2'')+(a\lambda\mu^2+b\mu^2)\phi''=0, \quad . \quad . \quad . \quad (26a)$$

$$\cos 2\alpha(\psi_1''+\psi_2'')+2\lambda\mu\phi''=0. \quad . \quad . \quad (27\,a)$$

So long as λ and μ are real, that is, so long as

$$\sin \alpha < \mathrm{V}_n/\mathrm{V}_{\mathrm{D}},$$

the above equations give quite reasonable results, with ψ_1 and ϕ representing real plane waves.

But when $\sin \alpha=\mathrm{V}_n/\mathrm{V}_{\mathrm{D}}$ we find that $\lambda=0$, and consequently $(27a)$, gives

$$\psi_2''=-\psi_1'', \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28a)$$

which shows that the rotation wave ψ_1 is reflected with changed sign, but with the same energy. Then $(26a)$ gives ϕ'', which is not zero, as might have been expected from the fact that the other reflected wave seems to carry all the energy. It must mean, however, that ϕ adds nothing to energy. The explanation is that the breadth of the ϕ wave, corresponding to a finite width of the ψ_1 and ψ_2 waves, is infinitesimal since a figure shows, in the general case, that the ratio of corresponding widths is $\lambda : \cos \alpha$. Thus, while ϕ'' is not small, it nevertheless occupies such a small volume of the material that its energy is infinitesimal.

Equation $(28a)$ indicates that, at this particular angle of incidence, the wave ψ_1 is totally reflected just as light waves can be totally reflected.

When $\sin \alpha>\mathrm{V}_n/\mathrm{V}_{\mathrm{D}}$ there is still total reflection, as we shall show.

The ϕ function we have assumed is still a solution of the differential equation for $\dot{\phi}$ so long as $(22a)$ is true, even if $\mu>1$; but it no longer represents a plane wave.

When $\mu>1$ we find that

$$\lambda^2=1-\mu^2=\text{negative},$$

$$=-\lambda_1{}^2 \text{ say},$$

i. e.,

$$\lambda=i\lambda_1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29a)$$

We get a real dilatation function if we now take

$$\phi=\tfrac{1}{2}\phi(\mathrm{V}_{\mathrm{D}}t-i\lambda_1x+\mu y),$$

$$+\tfrac{1}{2}\phi(\mathrm{V}_{\mathrm{D}}t+i\lambda_1x+\mu y). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30\,a)$$

The actual boundary condition that gave rise to $(27a)$ was

$$2\frac{\partial^2\phi}{\partial x\,\partial y}=\cos 2\alpha(\psi_1''+\psi_2'') \text{ at } x=0. \quad . \quad . \quad . \quad . \quad . \quad (31a)$$

When the value of ϕ in $(30a)$ is substituted in this it clearly gives

$$0=\psi_1''+\psi_2'',$$

again denoting total reflexion of the ψ_1 wave.

Also, merely by putting $-\lambda_1{}^2$ for λ^2 in $(26a)$, that equation gives correctly the value of $\phi''(\mathrm{V}_{\mathrm{D}}t+\mu y)$ even when $\mu>1$.

Thus a rotation wave follows the laws of optics. The only difference is that the refracted wave ϕ is refracted back into the material instead of out of the material. The law for total reflexion is precisely the same as in optics.

It is useful to observe here that ψ_2 alone is capable of carrying all the energy which ψ_1 brings to the boundary, and yet ψ_2 and ϕ clearly carry no more. This shows that the energy due to two waves is not the sum of the energies due to each taken separately.

Although it may not be easy to picture exactly what happens when any wave which is not a plane wave meets a free boundary, yet any wave whatever can be regarded as the sum of a number of spherical waves originating at a number, maybe an infinite number, of points; and each spherical wave can be split up into very thin pencils which are, in effect, plane waves. The rules for the reflexion of these pencils are the same as those for plane waves.

It now follows that any elastic disturbance travels along a rod, thick or thin, by waves which mainly strike the sides of the rod and proceed by a series of reflexions along broken lines going from side to side of the rod. The dilatation wave, every time it strikes the boundary, sends off a reflected dilatation wave of different intensity and the angle of reflexion is the same as for light rays; it also starts a rotation wave at each reflexion, the direction of motion of this wave being nearer the normal to the surface than the direction of the dilatation waves. Consequently, although rotation waves are being constantly generated at the boundary surface, they are continually dropping behind the waves which generate them.

Of course the rotation waves themselves, when they strike a free surface, generate anew the two types of waves, and the dilatation waves thus generated by reflexion will race ahead of the generating rotation waves, as can easily be seen from the results just proved.

One thing, however, is certain, and that is, that no elastic disturbance can travel along a rod with a velocity greater than V_D, and no rotational disturbance can travel faster than V_n.

When a thin rod is set in motion by actions at one point, say one end, and then left to itself, the true state of the rod at any later instant could be found by an investigation of the waves created by the impulse or impulses applied to it. But it will be realized from the simple examples we have just worked out that the solution of such a problem, by taking account of all the reflected waves, could be a tremendously complicated operation. This, however, is the exact method and would tell us the precise state of the rod at any instant. Owing to the difficulties of the problem this method is not used, except in one or two very ideal problems. Instead of the exact process the usual method in dealing with these rods is to discard altogether the fundamental equations of elasticity and to substitute in their place simpler equations, involving fewer unknown variables, and based on the intuitions of some of the ablest investigators into these problems—a Bernoalli and an Euler for example. These simpler equations have no *raison d'être* unless they agree very well with results derived by more exact methods based on the fundamental equations. The way to test these intuitional equations is to solve the same problems by the fundamental equations and by the simpler equations, and to find to what extent they disagree, as disagree they must.

Very little has been done in this respect to compare results. Love ('Elasticity,' 4th Edition, Arts. 201, 202) shows that the results obtained by the two methods agree in essentials for a cylindrical rod vibrating either longitudinally or transversely provided that the wave-length is very big compared with the radius of the rod. He did not, however, carry his investigations to shorter wave-lengths. Later in this paper I shall investigate in detail one particular case of a vibrating rod and shall prove the interesting result that, at a particular wave velocity a little below V_n, the vibrations, whether longitudinal or transverse, change their character very greatly and become *skin waves* of infinitely short wave-length. These are the same waves as Rayleigh investigated under the name of *surface waves* in an infinite solid—now regarded as having importance in earthquake tremors. All wave-lengths were possible in Rayleigh's case because his solid was regarded as infinite; but in a finite body the wave-length of skin waves must be small in comparison with the dimensions of the solid—the smallest dimension would be more accurate.

Longitudinal Vibrations.

Before going on to exact solutions of thin rod problems we shall briefly review the facts about the current simplified theories.

By means of the assumption that it is possible for the transverse tensile stresses P_y and P_z to be zero, or negligible, in a thin rod oscillating longitudinally in the x direction, it follows from the fundamental equation

$$E\frac{\partial u}{\partial x}=P_x-\sigma(P_y+P_z), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

that

$$P_x=E\frac{\partial u}{\partial x}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

By means of this the differential equation for the longitudinal motion of a uniform thin rod takes the form

$$V_E{}^2\frac{\partial^2 u}{\partial x^2}=\frac{\partial^2 u}{\partial t^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

This is precisely the same as the equation for plane waves in the general theory, but with a velocity of propagation V_E and not V_D or V_n. The general solution is

$$u=f(x-V_Et)+F(x+V_Et), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

which represents two wave forms travelling in opposite directions each with velocity V_E.

The above is the complete solution of the problem of the free longitudinal motion of a thin rod on the present assumptions and, for any particular case, all that is necessary is to find the functions f and F that fit the initial conditions, these conditions being, in their simplest form,

$$\left.\begin{aligned} u&=2\phi(x), \\ \frac{\partial u}{\partial t}&=2\psi'(x), \end{aligned}\right\}, \qquad (35)$$

when $t=0$. From these we find

$$\mathrm{F}(x)=\phi(x)+\frac{1}{\mathrm{V_E}}\psi(x), \qquad (36)$$

$$f(x)=\phi(x)-\frac{1}{\mathrm{V_E}}\psi(x). \qquad (37)$$

For an infinite rod the problem is completely solved by the substitution of these values for f and F in equation (34). If the rod is finite the problem is complicated by the successive reflexions from the ends, the result of which is that, as time goes on, more and more waves are travelling in both directions along the rod.

Another method of solving the same problem, and one which avoids the complication of reflected waves, is to write down particular solutions of (33) and to sum them, this summation being justified because the differential equation is linear. Thus

$$u=\{\mathrm{B}\sin cx+\mathrm{G}\cos cx\}\sin(c\mathrm{V_E}t+\alpha) \qquad (38)$$

is a solution of (33) for all values of B, G, c, α. Now Fourier showed, over a hundred years ago, that any arbitrary function of x which could possibly occur in the problem of a rod vibrating longitudinally, including any discontinuities due to a start from rest, can be represented between $x=0$ and $x=l$ by the function

$$\sum_{n=1}^{n=\infty}\mathrm{H}_n\sin\frac{n\pi x}{l}, \qquad (39)$$

or by

$$\sum_{n=1}^{n=\infty}\mathrm{K}_n\cos\frac{n\pi x}{l}, \qquad (40)$$

and these are all the theorems necessary to satisfy any reasonable boundary and initial conditions of the vibrating rod. These conditions must be such that no energy is supplied to the rod at the ends after the initial impulses; otherwise it is not a free rod. It is obvious, of course, that a *free* rod, once it is set in motion, must vibrate, and this is what makes Fourier series suitable for the representation of the free motion.

Equation (33) has solutions of the form

$$u=\mathrm{N}e^{c(x-\mathrm{V_E}t)}$$

as well as the forms shown in (38). Yet these are not necessary to satisfy the conditions of free motion. What then do such solutions represent? They must represent the state of the rod while the impulses are being applied to it, and while, in fact, applied forces might prevent vibrations.

Transverse vibrations.

Let us now consider the simplified equation for the transverse motion of a thin rod. The usual equation is

$$K^2V_E{}^2\frac{\partial^4 w}{\partial x^4}=-\frac{\partial^2 w}{\partial t^2}, \quad \ldots \quad (41)$$

where K is the radius of gyration of the section of the rod about the neutral axis, the axis of no tensile stress.

If we attempt to get a solution of this of the type

$$w=f(x-Vt), \quad \ldots \quad (42)$$

we find

$$K^2V_E{}^2f^{\text{iv}}(x-Vt)=-V^2f''(x-Vt). \quad \ldots \quad (43)$$

When a single letter X is written for $(x-Vt)$ this gives

$$f(X)=A+BX+G\sin(cX+\alpha), \quad \ldots \quad (44)$$

where

$$c=\pm\frac{V}{KV_E}. \quad \ldots \quad (45)$$

We can omit A+BX since these terms do not give any stresses. Thus we find that wave motion of only one type is possible for a rod in transverse motion, and this can have any velocity V along the rod; but the wave-length λ is connected with the velocity V by Rayleigh's equation

$$c=\frac{2\pi}{\lambda}=\frac{V}{KV_E}. \quad \ldots \quad (46)$$

For comparison with later results we shall write this in the form

$$K^2c^2=\frac{V^2}{V_E{}^2}. \quad \ldots \quad (47)$$

Equation (41) is, of course, satisfied if V is imaginary. Putting

$$V=iU, \quad \ldots \quad (48)$$

and then

$$c=\pm\frac{iU}{KV_E}=ic_1 \text{ say}, \quad \ldots \quad (49)$$

we could now write the value of f in the form

$$\begin{aligned} f(X)&=Ge^{i(cX+\alpha)} \\ &=Ge^{-c_1x+ic_1Vt+i\alpha}. \quad \ldots \quad (50)\end{aligned}$$

When we pick out real and imaginary parts we find, from this, solutions of the type

$$w=(B\cosh c_1x+D\sinh c_1x)\sin(c_1Vt+\alpha). \quad \ldots \quad (51)$$

This last result does not represent a travelling wave at all since its velocity iU is not real.

Thus we find that transverse waves of a particular type only can travel along a thin rod and these can have any velocity whatever. There is, however, a relation between wave-length and velocity. Contrast this with longitudinal motion where all waves, of whatever type, travel with the same velocity V_E along the rod.

Equation (46) shows that waves with velocity V_E have a length

$$\lambda = 2\pi K.$$

For a cylindrical rod of radius a this is

$$\lambda = \pi a = 3{\cdot}14\,a, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

and for a flat strip of rectangular section and depth $2b$ it is

$$\lambda = \frac{2\pi b}{\sqrt{3}} = 3{\cdot}63\,b \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

Rotary Inertia.

After equation (41) had been in vogue for some time it was pointed out by someone that the rotary inertia of the rod had been ignored. A new term was introduced into the equation to take this into account, and the reasoning was as follows :—

Assuming, as in the Bernoulli equilibrium theory, that plane sections of the rod remain plane and perpendicular to the central axis while the rod is in transverse motion, the angular velocity of an element dx is

$$\frac{\partial}{\partial t}\left(\frac{\partial w}{\partial x}\right)$$

and the angular momentum of this length dx is

$$AK^2\,dx\frac{\partial}{\partial t}\left(\frac{\partial w}{\partial x}\right), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

where A denotes the area of the section. When this is taken into account and the same Bernoulli distribution of stress assumed as before, the differential equation becomes

$$V_E^2K^2\frac{\partial^4 w}{\partial x^4} = -\frac{\partial^2 w}{\partial t^2} + K^2\frac{\partial^4 w}{\partial x^2\,\partial t^2}. \quad . \quad . \quad . \quad . \quad . \quad (55)$$

Adopting a solution of the usual type

$$w = B\sin c(x - Vt + \alpha), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (56)$$

we find
$$V_E^2K^2c^4 = c^2V^2 + K^2c^4V^2,$$

whence
$$K^2c^2 = \frac{V^2}{V_E^2 - V^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (57)$$

For small values of V^2/V_E^2 this is approximately the same as (47); but, for larger values of V^2, we find that the wave-length $\lambda(\equiv 2\pi/c)$ vanishes when $V^2 = V_E^2$, whereas with the less exact differential equation it vanished only when $V^2 = \infty$. On account of the importance of the short waves as carriers of energy, the equation (55) is vastly more reasonable from a physical standpoint than (41).

New Hypotheses.

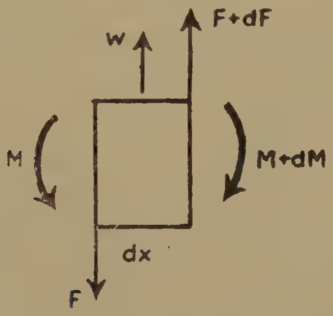
We shall now examine anew the problem of the transverse motion of a uniform rod without, at first, making any assumptions about the stress distribution. We shall, in fact, find how far we can go towards a solution of the problem without using the fundamental equations of elasticity and without making guesses, and we shall, on the way, indicate the flaws in the accepted equations.

As before, the x-axis is taken along the line of the undisturbed axis of the rod. A is the area of the cross-section AK^2, the moment of inertia (or second moment) of A about the neutral axis.

The displacements of the element $dx dA$ originally at x, y, z are u, v, w, the last being in the direction of the transverse motion. A line in the area A at x parallel to the y-axis is assumed to be an axis of principal moment of inertia of A.

We now consider the motion of the length dx of the rod under the action of bending moments and shear forces acting on the sections at its ends. These are M and F at section x.

Fig. 2.



The total momentum of this element in the z-direction is

$$dx\iint\rho\,dA\frac{\partial w}{\partial t}. \qquad (58)$$

The equation of translation is therefore

$$dF=\rho\,dx\iint\frac{\partial^2 w}{\partial t^2}dA. \qquad (59)$$

Now let us write

$$W\equiv\frac{1}{A}\iint w\,dA, \qquad (60)$$

this being a mean w for the section at x. This W is not the displacement of the central line but it will usually differ very little from that displacement.

From (60) we get

$$\frac{\partial^2 W}{\partial t^2}=\frac{1}{A}\iint\frac{\partial^2 w}{\partial t^2}dA,$$

and therefore (59) becomes

$$\frac{\partial F}{\partial x}=\rho A\frac{\partial^2 W}{\partial t^2}. \qquad (61)$$

We next deal with the rotational motion of the element of length dx.

The moment, about the neutral axis, of the momentum of the element $dx\,dA$ is

$$z\rho\,dA\,dx\frac{\partial u}{\partial t},$$

the moment being taken in the direction of $(M+dM)$. Consequently, the total moment of momentum of the element dx is

$$\rho\,dx\iint z\frac{\partial u}{\partial t}\,dA.$$

Let us now write

$$U\equiv\iint zu\,dA; \qquad \ldots\ldots\ldots \qquad (62)$$

and then the moment of momentum becomes

$$\rho\,dx\,A\frac{\partial U}{\partial t}.$$

Now the dynamical equation of moments for the element dx is

$$dM-F\,dx=\rho\,A\,dx\frac{\partial^2 U}{\partial t^2},$$

whence

$$\frac{\partial M}{\partial x}-F=\rho A\frac{\partial^2 U}{\partial t^2}. \qquad \ldots\ldots\ldots \qquad (63)$$

The equations (61) and (63) are exact, and they are all we can get out of dynamical equations without using the hypotheses of elasticity.

We can eliminate F from (61) and (63), the resulting equation being

$$\frac{\partial^2 M}{\partial x^2}=\rho A\left\{\frac{\partial^2 W}{\partial t^2}+\frac{\partial^2}{\partial t^2}\left(\frac{\partial U}{\partial x}\right)\right\}. \qquad \ldots\ldots \qquad (64)$$

By using the relation between shear stress and shear strain we can add another exact equation to (61) and (63). Thus

$$F=\iint n\left(\frac{\partial u}{\partial z}+\frac{\partial w}{\partial x}\right)dA$$

$$=nA\left\{S+\frac{\partial W}{\partial x}\right\}. \qquad \ldots\ldots\ldots \qquad (65)$$

where

$$S\equiv\frac{1}{A}\iint\frac{\partial u}{\partial z}\,dA. \qquad \ldots\ldots\ldots \qquad (66)$$

This is as far as we can get without using the fundamental equations of elasticity or making assumptions which are mere guesses.

Two assumptions have to be made to get the well-known equation (55) out of (64). These are

$$(\alpha) \qquad u=-z\frac{\partial W}{\partial x};$$

$$(\beta) \qquad P_x=E\frac{\partial u}{\partial x}.$$

These two give

$$\begin{aligned}M&=\iint zP_x\,dA\\&=\iint -Ez^2\frac{\partial^2 W}{\partial x^2}\,dA\\&=-EAK^2\frac{\partial^2 W}{\partial x^2}. \qquad (67)\end{aligned}$$

Also

$$\begin{aligned}U&=\frac{1}{A}\iint -z^2\frac{\partial W}{\partial x}\,dA\\&=-K^2\frac{\partial W}{\partial x}. \qquad (68)\end{aligned}$$

By means of these results equation (64) becomes

$$V_E{}^2K^2\frac{\partial^4 W}{\partial x^4}=-\frac{\partial^2 W}{\partial t^2}+K^2\frac{\partial^4 W}{\partial t^2\,\partial x^2}, \qquad (69)$$

which is the same as (55) but with W for w.

The assumptions (α) and (β), being mere guesses based only on somebody's intuition, are not likely to be exact. There is no way of finding out how near they are to the truth except by starting from the fundamental equations of elasticity and solving, by strictly mathematical methods, some problems of vibration of thin rods. Love, in his 'Elasticity' (4th edition, Art. 202), shows that, for large wave-lengths, the equation (47), namely

$$K^2c^2=V^2/V_E{}^2,$$

is approximately true and is true in the limit when $\lambda\to\infty$, in the case of a cylindrical rod; but as we have remarked before, he did not find how K^2c^2 varies with V^2 for larger values of V^2.

One result that follows from the assumption (α) shows with what care it is necessary to use these intuitional assumptions. Thus, using (α), we find

$$S=\frac{1}{A}\iint\frac{\partial u}{\partial z}\,dA=-\frac{\partial W}{\partial x}. \qquad (70)$$

Then (65) gives

$$F=0, \qquad (71)$$

and now (61) yields

$$\frac{\partial^2 W}{\partial t^2}=0. \qquad (72)$$

Thus, by using (α) in the wrong place, and combining it with exact equations, we are led to an obvious absurdity.

By a better choice of guesses than (α) and (β) it is possible to arrive at a better differential equation than (55), as we shall now show.

On eliminating F from (61) and (65) we get

$$\frac{1}{V_n{}^2}\frac{\partial^2 W}{\partial t^2}-\frac{\partial^2 W}{\partial x^2}=\frac{\partial S}{\partial x}. \qquad (73)$$

Further, we have previously found that, if

$$P_y=0, \quad P_z=0, \qquad (74)$$

then

$$P_x=E\frac{\partial u}{\partial x}, \qquad (75)$$

which is the assumption (β). Now (74) would be true for a rod at rest under transverse forces, and could therefore be expected to be nearly true for a rod in which the accelerations are small, that is, for vibrations of long wave-lengths.

If, instead of (74), we had assumed that

$$\frac{\partial v}{\partial y}=0, \quad \frac{\partial w}{\partial z}=0, \qquad (76)$$

we should have found

$$E\frac{\partial u}{\partial x}=P_x\left\{1-\frac{2\sigma^2}{1-\sigma}\right\}, \qquad (77)$$

which becomes, in terms of the constants m and n,

$$P_x=(m+n)\frac{\partial u}{\partial x}. \qquad (78)$$

The assumptions (74) and (76) may be regarded as opposite extremes. In the former the full effect of the Poisson contraction due to P_x has time to take place, and in the latter no effect at all takes place. This latter condition would be approached very probably for quick vibrations and the former for slow vibrations. For most vibrations the truth probably lies between the two sets of assumptions.

We may reasonably take then

$$P_x=f\frac{\partial u}{\partial x}. \qquad (79)$$

where f is really a function of V. which is almost certainly between E and $(m+n)$.

By means of (79) equation (64) becomes

$$V_f{}^2\frac{\partial^2 U'}{\partial x^2}-\frac{\partial^2 U'}{\partial t^2}=\frac{\partial^2 W}{\partial t^2}, \qquad (80)$$

where U′ is written for $\frac{\partial U}{\partial x}$ and V_f has an obvious meaning.

Our second assumption is that

$$u=z\mu, \qquad (81)$$

where μ is a function of x and t only. This still makes the plane sections remain plane throughout the motion, but does not necessarily force them to be perpendicular to the axis of the rod. In other words, the shear force F is not constrained to be zero.

Thus we get

$$S=\frac{1}{A}\iint\frac{\partial u}{\partial z}dA=\mu, \qquad (82)$$

$$U=\frac{1}{A}\iint z u\,dA=K^2\mu. \qquad (83)$$

Equation (80) now takes the form

$$V_f^2 \frac{\partial^2 \mu'}{\partial x^2} - \frac{\partial^2 \mu'}{\partial t^2} = \frac{\partial^2 W}{\partial t^2} \cdot \frac{1}{K^2}, \tag{84}$$

and (73) the form

$$\frac{1}{V_n^2} \frac{\partial^2 W}{\partial t^2} - \frac{\partial^2 W}{\partial x^2} = \mu', \tag{85}$$

μ' being $\frac{\partial \mu}{\partial x}$.

The elimination of μ' from these gives

$$\left\{ V_f^2 \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial t^2} \right\} \left\{ \frac{1}{V_n^2} \frac{\partial^2 W}{\partial t^2} - \frac{\partial^2 W}{\partial x^2} \right\} = \frac{1}{K^2} \frac{\partial^2 W}{\partial t^2}. \tag{86}$$

This, then, is almost as good as we can hope to get by this kind of method.

The assumption (81) could be slightly improved because, as it stands, it gives

$$\frac{\partial u}{\partial z} = \mu, \tag{87}$$

and thus makes this, a term in the expression for S_y, a constant across the section. But we know that the shear stress vanishes at the upper and lower edges of the section and so a value of $\frac{\partial u}{\partial z}$, which has a maximum at the centroid of the section and becomes small or vanishes at the limits, would be better. For a rectangular section of depth $2b$ we might take

$$\frac{\partial u}{\partial z} = \mu \frac{b^2 - z^2}{b^2}, \tag{88}$$

or

$$\frac{\partial u}{\partial z} = \mu \cos \frac{\pi z}{2b}, \tag{89}$$

and expect to get better results. The second of these assumptions gives

$$S = \frac{2\mu}{\pi} = \xi \text{ say,}$$

$$U = \frac{8}{\pi^3} b^2 \mu = \frac{12}{\pi^2} K^2 \xi.$$

Then, instead of (86), we get

$$\left\{ V_f^2 \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial t^2} \right\} \left\{ \frac{1}{V_n^2} \frac{\partial^2 W}{\partial t^2} - \frac{\partial^2 W}{\partial x^2} \right\} = \frac{\pi^2}{12} \cdot \frac{1}{K^2} \frac{\partial^2 W}{\partial t^2}{}^{*} \tag{90}$$

There is not much to choose between (86) and (90). The difference between them is in terms that are small except for waves with big velocities, *i. e.* velocities of the order V_n or more. It will be found later that they are both much more accurate equations than the usual equation (55), which is supposed to take account of rotary inertia.

---

[*] S. Timoshenko gets a result of the same form as this, by rather different reasoning, in his book ‘Vibration Problems in Engineering,’ Art. 41.

Once more let us find what function of V^2 the quantity K^2c^2 is if (86) is true.

The substitution

$$W=B\sin c(x-Vt+\alpha)$$

in (86) gives

$$K^2c^2=\frac{V_n{}^2V^2}{(V_f{}^2-V^2)(V_n{}^2-V^2)}, \quad \ldots \ldots \quad (91)$$

which becomes, for small $V^2/V_n{}^2$,

$$K^2c^2=\frac{V^2}{V_f{}^2}, \quad \ldots \ldots \quad (92)$$

and we decided that V_f should be V_E for small V. Also c^2 approaches ∞ and therefore λ^2 approaches zero as V^2 approaches $V_n{}^2$. That is, λ varies from ∞ to zero as V varies from 0 to V_n. This is physically more reasonable than the conclusion from (57) that $\lambda=0$ only when $V=V_E$.

Vibrations of a Flat Strip or Broad Beam. Exact Equations.

The question we shall now examine in detail is the transverse motion of a broad beam of depth $2b$ and relatively large width, something of the dimensions of a steel rule. For simplicity of analysis we shall assume that there is no displacement v, thus treating the problem as a two-dimensional one. In our treatment the beam is, in fact, a strip of an infinitely wide plate of uniform thickness. One important difference between a plate and a beam, even in the ordinary simplified theories, is that the constant E in the beam theory is replaced by the constant $E/(1-\sigma^2)$ in the plate theory. If our reasoning up to now is right this means that the equation connecting c and V for a plate should be, for small V/V_E,

$$K^2c^2=(1-\sigma^2)\frac{V^2}{V_E{}^2}, \quad \ldots \ldots \quad (93)$$

whereas it was

$$K^2c^2=\frac{V^2}{V_E{}^2}$$

for a thin rod. This we shall find to be true.

Instead of using a sine function for the typical wave we shall find it more convenient to use the form $e^{c(x-Vt)}$, so as to avoid imaginaries in the working. We can always replace c by ic_1 when necessary.

The two displacements are u and w and they are got from one dilatation wave and one rotation wave.

For dilatation

$$u=\frac{\partial\phi}{\partial x}, \quad w=\frac{\partial\phi}{\partial z}, \quad \ldots \ldots \quad (94)$$

where ϕ is a function satisfying

$$V_D{}^2\nabla^2\phi=\frac{\partial^2\phi}{\partial t^2}. \quad \ldots \ldots \quad (95)$$

The rotation function is ψ and it satisfies the equation

$$V_n^2\nabla^2\psi=\frac{\partial^2\psi}{\partial t^2}. \quad (96)$$

The corresponding displacements are

$$u=-\frac{\partial\psi}{\partial z}, \quad w=\frac{\partial\psi}{\partial x}. \quad (97)$$

These latter displacements satisfy the condition that

$$\Delta=\frac{\partial u}{\partial x}+\frac{\partial w}{\partial z}=0, \quad (98)$$

and also satisfy the differential equations for the rotational motion.

Now we assume that all our displacements have a factor $e^{c(x-Vt)}$, which we shall sometimes denote briefly by X.

Putting, in equation (95),

$$\phi=\phi_1 e^{c(x-Vt)}, \quad (99)$$

and in (96)

$$\psi=\psi_1 e^{c(x-Vt)}, \quad (100)$$

we find

$$\frac{d^2\phi_1}{dz^2}=-c^2h^2\phi_1,$$

$$\frac{d^2\psi_1}{dz^2}=-c^2k^2\psi_1,$$

where

$$h^2=1-\frac{V^2}{V_D^2}, \quad (101)$$

$$k^2=1-\frac{V^2}{V_n^2}. \quad (102)$$

For transverse motion u should be zero on the x-axis and w should be a maximum there. So we take

$$c\phi_1=T\sin chz,$$

i. e.,

$$c\phi=Te^{c(x-Vt)}\sin chz.$$

The corresponding displacements are

$$\left.\begin{aligned}u&=Te^{c(x-Vt)}\sin chz,\\ w&=hTe^{c(x-Vt)}\cos chz,\end{aligned}\right\} \quad (103)$$

For the rotation we find the displacements

$$\left.\begin{aligned}u&=kHe^{c(x-Vt)}\sin ckz,\\ w&=He^{c(x-Vt)}\cos ckz.\end{aligned}\right\} \quad (104)$$

Finally the total displacements of the right type are

$$u=e^{c(x-Vt)}\{T\sin chz+kH\sin ckz\}, \quad (105)$$

$$w=e^{c(x-Vt)}\{hT\cos chz+H\cos ckz\}. \quad (106)$$

The dilation is

$$\Delta=\frac{\partial u}{\partial x}+\frac{\partial w}{\partial z}$$

$$=c\frac{V^2}{V_D^2}Te^{c(x-Vt)}\sin chz. \quad (107)$$

The two stresses that must vanish at the upper and lower surfaces of the strip are P_z and S_y. Now

$$P_z=(m-n)\Delta+2n\frac{\partial w}{\partial z}.$$

So
$$\frac{P_z}{2nX}=\frac{m-n}{2n}c\frac{V^2}{V_D^2}T\sin chz-ch^2T\sin chz-ckH\sin ckz$$
$$=\left\{\frac{m+n}{2n}\frac{V^2}{V_D^2}-1\right\}cT\sin chz-ckH\sin ckz. \quad . \quad . \quad . \qquad (108)$$

This vanishes at $z=\pm b$. Therefore, since

$$nV_D^2=(m+n)V_n^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (109)$$

the boundary condition is

$$T(1+k^2)\sin chb=-2kH\sin ckb. \quad . \quad . \quad . \quad . \quad . \qquad (110)$$

The shear stress S_y is given by

$$\frac{S_y}{ncX}=\frac{\partial u}{\partial z}+\frac{\partial w}{\partial x}$$
$$=hT\cos chz+k^2H\cos ckz$$
$$+hT\cos chz+H\cos ckz$$
$$=2hT\cos chz+(1+k^2)H\cos ckz. \quad . \quad . \quad . \quad . \quad . \qquad (111)$$

The condition that S_y is zero at $z=\pm b$ gives

$$2hT\cos chb=-(1+k^2)H\cos ckb. \quad . \quad . \quad . \quad . \quad . \qquad (112)$$

From the two boundary conditions (109) and (112) the ratio T/H can be eliminated, the result being

$$\frac{\tan ckb}{\tan chb}=\frac{(1+k^2)^2}{4hk}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (113)$$

This equation gives a relation between cb and V, since h^2 and k^2 are mere functions of V^2. It corresponds to our previous equations connecting c^2K^2 and V^2. We can make comparisons with our previous equations, such as (91) for example, only by laboriously finding from (113) values of cb corresponding to a series of values of V^2.

It is worth while to show that (113) is really an equation connecting c^2b^2 and V^2; that is, odd powers of c really do not occur. Thus

$$\frac{\dfrac{\tan ckb}{ckb}}{\dfrac{\tan chb}{chb}}=\frac{(1+k^2)^2}{4k^2}=\frac{\left(2-\dfrac{V^2}{V_n^2}\right)^2}{4k^2}$$

$$=1+\frac{V^4}{4k^2V_n^4}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (114)$$

Now $\tan\theta/\theta$ is an even function of θ, and consequently the left-hand side of (114) involves no odd powers of cb. The equation then can be regarded as expressing c^2b^2, for which we shall write R^2, as a function of V^2. This relation can be shown in a graph and we shall see that this graph has many separate branches. The most important branch for this problem is the one that passes through the origin, where R^2 and V^2 are both zero. This one branch, in fact, corresponds to the whole graph represented by (91) between $V^2=-\infty$ and $V^2=V_n^2$.

For the present we will investigate the branch of the curve connecting R^2 and V^2/V_n^2 which passes through the origin.

Since

$$\frac{\tan\theta}{\theta}=1+\frac{1}{3}\theta^2+\frac{2}{15}\theta^4+\ldots, \quad . \quad . \quad . \quad . \quad (115)$$

it is clear that, for small values of θ at least, this function, which we shall denote by $F(\theta)$, increases as θ increases. Actually the function goes on increasing up to $\theta=\pi/2$, although the series is convergent only up to $\pi/4$. Now for real values of V less than V_n, *i. e.*, for positive values of V^2 less than V_n^2, k^2 is less than h^2, and consequently $F(ckb)$ is less than $F(chb)$. Equation (114) has therefore no real roots for R in this region. But if we put ic_1 for c and iR_1 for R (R_1 being now $2\pi b/\lambda$), then

$$F(i\theta)=\frac{i\tanh\theta}{i\theta}=\frac{\tanh\theta}{\theta}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (116)$$

and so (114) becomes

$$\frac{\dfrac{\tanh kR_1}{kR_1}}{\dfrac{\tanh hR_1}{hR_1}}=1+\frac{V^4}{4k^2V_n^4}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (117)$$

Graphs of $\tanh\theta$ and θ will show at once that $F(i\theta)$ is a decreasing function of θ ranging from the value 1 at $\theta=0$ to the value 0 at $\theta=\infty$. It is now clear that (117) has real roots for R_1^2 for values of V^2 varying from small values up to some limit which we can find. Let us first find the value of R_1^2 in terms of a small value of V^2/V_n^2.

Using the expansion (115) in either (114) or (117) we get

$$\frac{1-\frac{1}{3}k^2R_1^2+}{1-\frac{1}{3}h^2R_1^2+}=1+\frac{V^4}{4k^2V^4},$$

i. e.,

$$1+\tfrac{1}{3}(h^2-k^2)R_1^2+=1+\frac{V^4}{4k^2V_n^4},$$

from which

$$\frac{1}{3}\frac{V^2}{V_n^2}\left\{1-\frac{V_n^2}{V_D^2}\right\}R_1^2+=\frac{V^4}{4k^2V_n^4}.$$

So, as far as the order V^2/V_n^2,

$$\frac{1}{3}R_1^2\left\{1-\frac{V_n^2}{V_D^2}\right\}=\frac{1}{4}\frac{V^2}{V_n^2},$$

or

$$\frac{1}{3}R_1^2=\frac{1}{4}\frac{V^2}{V_n^2}\cdot\frac{V_D^2}{V_D^2-V_n^2}=\frac{1}{4}\frac{V^2}{V_E^2}\cdot\frac{V_E^2}{V_n^2}\cdot\frac{V_D^2}{V_D^2-V_n^2}.$$

Now

$$\frac{V_D^2}{V_n^2}=\frac{m+n}{n},$$

$$\frac{V_E^2}{V_n^2}=\frac{E}{n}.$$

So
$$\frac{V_E^2}{V_n^2}\cdot\frac{V_D^2}{V_D^2-V_n^2}=\frac{E}{n}\cdot\frac{m+n}{m},$$
$$=2(1+\sigma)\times 2(1-\sigma),$$
$$=4(1-\sigma^2), \qquad (118)$$

and therefore
$$K^2c^2=\tfrac{1}{3}R_1^2=(1-\sigma^2)\frac{V^2}{V_E^2}, \qquad (119)$$

K being here written for the radius of gyration of the cross-section of the flat strip which we are regarding as a thin rod. As we mentioned before, this result is what we should expect for a broad strip, which is, in effect, a thin plate, for in the equations of a plate $E/(1-\sigma^2)$ takes the place of E in a beam. So far then, the usual plate theory agrees perfectly with the result obtained from our more exact equations.

To trace the further variation of R_1^2 we put (117) in the form
$$\frac{\tanh kR_1}{\tanh hR_1}=\frac{k}{h}\left(1+\frac{V^4}{4k^2V_n^4}\right)$$
$$=\frac{(1+k^2)^2}{4hk}. \qquad (120)$$

Here we see that, when $V^2=V_n^2$ and therefore $k=0$, the right-hand side of (120) becomes infinite. This same right-hand side starts at 1 when $V^2=0$ and increases to ∞ at $V^2=V_n^2$, but it does not continually increase in this region because, as we shall show, it comes back to the value unity for a particular value of V^2/V_n^2.

It is clear that $R_1=\infty$ is possible if the equation
$$\frac{(1+k^2)^2}{4hk}=1 \qquad (121)$$
has a real root between $V^2=0$ and $V^2=V_n^2$. It can be shown arithmetically that it has a real root for possible values of σ. When $\sigma=0{\cdot}3$, for instance,
$$h^2=1-\frac{V^2}{V_D^2}=1-\frac{n}{m+n}\frac{V^2}{V_n^2}=1-\frac{2}{7}\frac{V^2}{V_n^2}. \qquad (122)$$

So equation (121) becomes, when squared,
$$\left(2-\frac{V^2}{V_n^2}\right)^4=16\left(1-\frac{2}{7}\frac{V^2}{V_n^2}\right)\left(1-\frac{V^2}{V_n^2}\right). \qquad (123)$$

It can be verified that this is satisfied by
$$V^2/V_n^2=0{\cdot}860. \qquad (124)$$

Thus is appears that R_1^2 varies from 0 to ∞ while V^2 varies from 0 to $0{\cdot}86\ V_n^2$. In order to find enough corresponding values of R_1^2 and V^2/V_n^2 between these extreme values of V^2, it is necessary to choose values of V^2/V_n^2 and use trial and error methods to find R_1 from (120). It seemed easier to me, with the tables available, to write l for the right-hand side of (120) and transform the equation thus
$$\frac{\sinh\,(h+k)R_1}{\sinh\,(h-k)R_1}=\frac{l+1}{l-1}. \qquad (125)$$

There is no need to give the details of the calculations. We give only a table of results, which can be verified by means of mathematical tables.

TABLE A.

V^2/V_n^2	0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·86
R_1^2	0	0·120	0·288	0·510	0·85	1·38	2·38	4·55	12·0	∞
R_1	0	0·346	0·537	0·712	0·92	1·17	1·55	2·135	3·46	∞

The tangent at the origin to the curve obtained when R_1^2 is plotted against V^2/V_n^2 is given by (119), and its slope is

$$\frac{dR_1^2}{d(V^2/V_n^2)}=3(1-\sigma^2)\frac{V_n^2}{V_E^2}$$

$$=3\frac{1-\sigma^2}{2(1+\sigma)}=\frac{3}{2}(1-\sigma)=\frac{21}{20}(\sigma=0\cdot3). \quad . \quad . \quad (126)$$

It will be noticed, on looking at the above table, that the curve connecting R_1^2 and V^2/V_n^2 very quickly deviates from the tangent, for, following the tangent, the table would be

TABLE A_1.

V^2/V_n^2	0	0·1	0·2	0·3
R_1^2	0	0·105	0·210	0·315

Thus the linear rule has an error of 12 per cent. in R_1^2 and 6 per cent. in λ at $V^2=0\cdot1\,V_n^2$, and this corresponds to a wave-length

$$\lambda=\frac{2\pi b}{c_1 b}=9\cdot08\times 2b, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (127)$$

that is, over nine times the total depth of the strip. This gives some idea of the limits of accuracy of the usual differential equation for vibrating beams, since this usual equation gives the linear law of Table A_1 and the error will be very much the same for all beams.

The numbers in Table A are shown in a graph (fig. 3) giving $R^2=-R_1^2$ against V^2/V_n^2. The tangent at the origin is also shown, this line being the graph given by the usual differential equation (41).

The branch of the curve connecting R^2 and V^2/V_n^2, which passes through the origin, is not completely represented by Table A, for there are possible imaginary values of V^2 on this same branch. When we put

$$V^2=-U^2,$$

we find

$$k^2=1+\frac{U^2}{V_n^2}, \quad h^2=1+\frac{2}{7}\frac{U^2}{V_n^2} \quad (\sigma=0\cdot3), \quad . \quad . \quad . \quad (128)$$

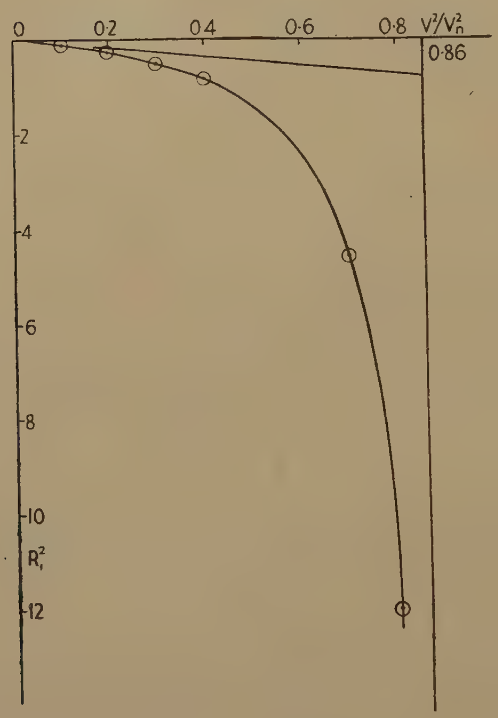
and consequently $k^2>h^2$ in this region, where U is real and V imaginary It follows, therefore, that (114) can be satisfied by real values of cb.

These real values of cb give rise to terms corresponding to the usual $\cosh cx$ and $\sinh cx$ terms in the ordinary theory of vibrating rods, whereas the imaginary values of cb, which we have just finished calculating, give rise to terms $\cos cx$ and $\sin cx$, each term having also a time factor of course.

For real values of $cb(\equiv R)$ then we go back to (113), namely,

$$\frac{\tan kR}{\tan hR}=\frac{(1+k^2)^2}{4hk}=\frac{k}{h}\left\{1+\frac{U^4}{4k^2V_n^4}\right\}. \quad \text{(129)}$$

Fig. 3.



This, of course, starts with

$$\tfrac{1}{3}R^2=-(1-\sigma^2)\frac{V^2}{V_E^2}=+(1-\sigma^2)\frac{U^2}{V_E^2}, \quad \text{(130)}$$

as in (119), for small values of U^2/V_n^2; and when $U\to\infty$

$$\frac{\tan kR}{\tan hR}\to\sqrt{\frac{7}{2}}\,\frac{U^4}{4U^2V_n^2}\to\infty; \quad \text{(130 a)}$$

so

$$kR\to\frac{\pi}{2},$$

or

$$\frac{U^2}{V_n^2}R^2\to\frac{\pi^2}{4}. \quad \text{(131)}$$

This obviously means that R is zero when U is infinite but the product UR is finite. We shall need this result later.

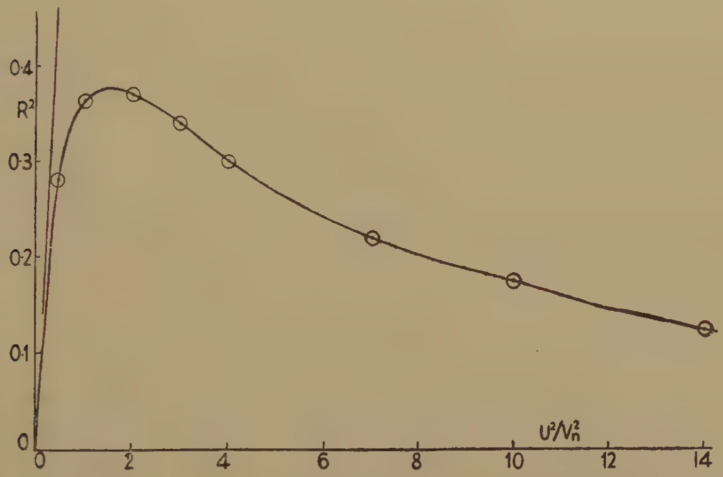
Thus R^2 varies from 0 to 0 as U^2 varies from 0 to ∞, and R^2 is positive for intermediate values since kR varies from 0 to $\pi/2$. Consequently R has at least one maximum in this region.

After laborious arithmetic the following table has been compiled for this branch of the curve:—

TABLE B.

$U^2/V_n{}^2$	0	0·5	1	2	3	4	7	10	14	∞
R^2	0	0·281	0·365	0·371	0·341	0·300	0·222	·1767	0·1448	0

Fig. 4.



These are plotted on a separate graph in fig. 4, because the nature of this curve does not show up very well when it is plotted on the same scale as the other part of the curve, derived from Table A. The tangent at the origin is again shown for the sake of comparison, with the results derived from the usual equation (41). The tangent, in fact, is the graph of equation (47) except that $E'=E/(1-\sigma^2)$ takes the place of E because our flat strip is a plate and not strictly a thin rod. This tangent is, of course, a continuation of the tangent in fig. 3.

Variation of the Stresses.

In equations (111) and (108) we found

$$\frac{S_y}{ncX}=2hT\cos chz+(1+k^2)H\cos ckz\;; \quad . \quad . \quad . \quad . \quad . \quad (140)$$

$$\frac{P_z}{ncX}=-(1+k^2)T\sin chz-2kH\sin ckz\;; \quad . \quad . \quad . \quad . \quad . \quad (141)$$

to which we add

$$P_x=(m-n)\Delta+2n\frac{\partial u}{\partial x},$$

whence

$$\frac{P_x}{ncX}=\left\{2+\frac{m-n}{m+n}\frac{V^2}{V_n}\right\}T\sin chz+2kH\sin ckz. \quad . \quad . \quad (142)$$

Now the boundary conditions $S_y=0$ at $z=\pm b$ gave

$$\begin{aligned}(1+k^2)H\cos ckb&=-2hT\cos chb,\\&=L \text{ say}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (143)\end{aligned}$$

Therefore we can write

$$\frac{S_y}{ncXL}=\frac{\cos ckz}{\cos ckb}-\frac{\cos chz}{\cos chb}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (144)$$

$$\frac{P_x}{ncXL}=\frac{2k}{1+k^2}\frac{\sin ckz}{\cos ckb}-\left\{1+\frac{m-n}{m+n}\cdot\frac{V^2}{2V_n^2}\right\}\frac{1}{h}\frac{\sin chz}{\cos chb}. \quad . \quad (145)$$

By using the other boundary condition we can express P_z in the form

$$\frac{P_z}{ncXN}=\frac{\sin ckz}{\sin ckb}-\frac{\sin chz}{\sin chb}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (146)$$

where $$N=(1+k^2)T\sin chb=-2kH\sin ckb.$$

The total shear force F across unit width of the section, and the bending moment M on this same unit width are given by

$$\begin{aligned}\frac{F}{2nbcXL}&=\frac{1}{2b}\int_{-b}^{b}\frac{S_n}{ncXL}dz\\&=\frac{\tan ckb}{ckb}-\frac{\tan chb}{chb}; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (147)\end{aligned}$$

and

$$\begin{aligned}\frac{M}{2nb^2cXL}&=\frac{1}{2b^2}\int_{-b}^{b}\frac{P_x z\,dz}{ncXL},\\&=\frac{2k}{1+k^2}\frac{1}{ckb}\left\{\frac{\tan ckb}{ckb}-1\right\}\\&\quad-\frac{1}{h}\left\{1+\frac{m-n}{m+n}\frac{V^2}{2V_n^2}\right\}\frac{1}{chb}\left\{\frac{\tan chb}{chb}-1\right\} \quad . \quad . \quad (148)\end{aligned}$$

When c is imaginary, as it is when V^2 lies between 0 and $0{\cdot}86\ V_n^2$, we write ic_1 for c and get

$$\begin{aligned}X=e^{ic_1(x-Vt)}&=\cos c_1(x-Vt)+i\sin c_1(x-Vt)\\&=X_1+iX_2 \text{ say}.\end{aligned}$$

Now all the expressions for the stresses and for F and M above are analytically correct whether c is real or imaginary. But in order to get real results we have to pick out the real parts of the expressions for the stresses as well as for F and M. Thus

$$-\frac{S_y}{nc_1LX_2}=\frac{\cosh c_1kz}{\cosh c_1kb}-\frac{\cosh c_1hz}{\cosh c_1hb}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (149)$$

$$-\frac{P_x}{nc_1LX_1}=\frac{2k}{1+k^2}\frac{\sinh c_1kz}{\cosh c_1kb}-\left\{1+\frac{m-n}{m+n}\frac{V^2}{2V_n^2}\right\}\frac{1}{h}\frac{\sinh c_1hz}{\cosh c_1hb}. \qquad (150)$$

$$-\frac{F}{2nbc_1LX_2}=\frac{\tanh c_1kb}{c_1kb}-\frac{\tanh c_1hb}{c_1hb} \qquad (147\,a)$$

$$+\frac{M}{2nb^2c_1LX_1}=\frac{2k}{1+k^2}\frac{1}{c_1kb}\left\{\frac{\tanh c_1kb}{c_1kb}-1\right\}$$
$$-\frac{1}{h}\left\{1+\frac{m-n}{m+n}\frac{V^2}{2V_n^2}\right\}\frac{1}{c_1hb}\left\{\frac{\tanh c_1hb}{c_1hb}-1\right\}. \qquad (148\,a)$$

Here we have used

$$\sin i\theta=i\sinh\theta,\quad \cos i\theta=\cosh\theta.$$

Also,
$$\frac{P_z}{nc_1X_1L}=\frac{1+k^2}{2h}\frac{\sinh c_1hz}{\cosh c_1hb}-\frac{2k}{1+k^2}\frac{\sinh c_1kz}{\cosh c_1hb} \qquad (151)$$

It is more convenient to use L here rather than N because, if L is real then N is imaginary.

The significant stresses are S_y and P_x, and we want to trace their variation as V^2 increases from 0 to $0{\cdot}86\,V_n^2$.

When V^2 is very small, and therefore, as we know from our calculations, c_1 also very small, we find

$$\frac{S_y}{nc_1X_2L}=\frac{1+\frac{1}{2}c_1^2k^2z^2}{1+\frac{1}{2}c_1^2k^2b^2}-\frac{1+\frac{1}{2}c_1^2h^2z^2}{1+\frac{1}{2}c_1^2h^2b^2} \text{ as far as } c_1^2,$$
$$=\tfrac{1}{2}c_1^2(h^2-k^2)(b^2-z^2)$$
$$=\tfrac{1}{2}c_1^2V^2\left(\frac{1}{V_n^2}-\frac{1}{V_D^2}\right)(b^2-z^2)=\frac{1}{2}\frac{m}{m+n}\frac{V^2}{V_n^2}c_1^2(b^2-z^2),$$
$$=G(b^2-z^2). \qquad (152)$$

Since c_1b^2 is proportional to V^2 it follows that G is proportional to V^4.

It is thus seen that the shear stress has a parabolic distribution across the section, and the mean stress is therefore two-thirds of the maximum, which occurs at $z=0$.

Let us now examine the variation of P_x when V is small.

When V is small we know that c_1b is of the same order as V. So, in expanding P_x, we must use, for safety,

$$\cos c_1kb=1-\tfrac{1}{2}c_1^2b^2 \text{ as far as } c_1^2;$$
$$\cos c_1hb=1-\tfrac{1}{2}c_1^2b^2.$$

Therefore

$$\frac{P_x}{nc_1X_1L}=\frac{2k^2c_1z}{(1+k^2)(1-\frac{1}{2}c_1^2b^2)}-\left\{1+\frac{m-n}{m+n}\frac{V^2}{2V_n^2}\right\}\frac{c_1z}{1-\frac{1}{2}c_1^2b^2}$$
$$=\frac{c_1z}{1-\frac{1}{2}c_1^2b^2}\left\{1-\frac{1}{2}\frac{V^2}{V_n^2}-1-\frac{m-n}{m+n}\frac{V^2}{2V_n^2}\right\}$$
$$=-\frac{m}{m+n}\frac{V^2}{V_n^2}c_1z.$$

Observe that

$$\frac{\max \mathrm{S}_y}{\max \mathrm{P}_x}=\tfrac{1}{2}c_1 b=\frac{\sqrt{3(1-\sigma^2)}}{2}\,\frac{\mathrm{V}}{\mathrm{V}_{\mathrm{E}}},$$

which is small since $\mathrm{V}/\mathrm{V}_{\mathrm{E}}$ is small.

So, for small values of V, the stress distributions across the section are exactly the same as are assumed in the simple theory of bending in equilibrium of a beam of rectangular section. This really justifies the usual differential equation for vibrations of a rod when V is small, that is, when the accelerations are small, and this, we found, meant when the wave-length is big compared with the depth.

Let us now consider the other extreme, where V^2 is nearly equal to $0{\cdot}86\ \mathrm{V}_n{}^2$ and c_1 consequently nearly infinite. In this case we may take, for large positive θ,

$$\cosh\theta=\tfrac{1}{2}e^{\theta},\quad \sinh\theta=\tfrac{1}{2}e^{\theta},$$

and therefore, except where z is so small that $c_1 z$ is not big,

$$\frac{2\mathrm{S}_y}{nc_1\mathrm{X}_2\mathrm{L}}=e^{-c_1k(b-z)}-e^{-c_1h(b-z)}\quad (z\text{ positive})\quad .\ \ .\ \ .\quad (154)$$

$$=e^{-c_1k(b+z)}-e^{-c_1h(b+z)}\quad (z\text{ negative}).\quad .\ \ .\quad (155)$$

In the first case $(b-z)$ is the distance from the nearest free surface and in the second case $(b+z)$ is the corresponding distance. Both results then take the form

$$\frac{2\mathrm{S}_y}{nc_1\mathrm{X}_2}=\mathrm{L}(e^{-c_1kd}-e^{-c_1hd}),\quad .\ \ .\ \ .\ \ .\ \ .\ \ .\ \ .\quad (156)$$

where d is the distance from the nearest boundary.

Now at the boundary, of course, $\mathrm{S}_y=0$, and at a short distance from the boundary both terms in the brackets in equation (156) are very small because c_1 is big and k and h are finite. In the short space where $c_1 d$ is a reasonable finite quantity (and there is such a space since $c_1 d=0$ at the surface and quickly becomes big), the stress S_y is a reasonably finite stress. Thus S_y is zero at the boundary, very suddenly shoots up to a maximum and then very suddenly drops to nearly zero as the point we follow across the section moves away from the boundary. The stress distribution is something like that shown in fig. 5, the stresses being indicated by ordinates perpendicular to the width $2b$.

The stress P_x does not vanish at the boundary and so its variations are somewhat different. With the same approximations as for S_y we get

$$\frac{\mathrm{P}_x}{nc_1\mathrm{X}_1\mathrm{L}}=\frac{2k}{1+k^2}e^{-c_1k(b-z)}-\left\{1+\frac{m-n}{m+n}\,\frac{\mathrm{V}^2}{2\mathrm{V}_n{}^2}\right\}\frac{1}{h}e^{-c_1h(b-z)}\quad (z\text{ positive}).\quad .\ \ .\ \ .\quad (157)$$

At $b=z$ this becomes

$$\frac{\mathrm{P}_x}{nc_1\mathrm{X}_1\mathrm{L}}=\frac{2k}{1+k^2}-\frac{1}{h}\left\{1+\frac{3}{14}\,\frac{\mathrm{V}^2}{\mathrm{V}_n{}^2}\right\}.\quad .\ \ .\ \ .\ \ .\ \ .\quad (158)$$

The rough numerical values near $V^2=0{\cdot}86\,V_n^2$ are

$$\left.\begin{aligned} k^2&=0{\cdot}14, & k&=0{\cdot}375,\\ h^2&=0{\cdot}754, & h&=0{\cdot}87.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (159)$$

So, roughly,

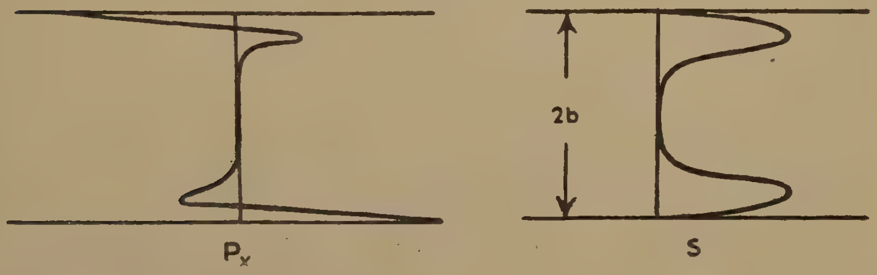
$$\begin{aligned}\frac{P_x}{nc_1X_1L}&=\frac{0{\cdot}75}{1{\cdot}14}-\frac{1}{0{\cdot}87}\left(1+\frac{3}{14}\times 0{\cdot}86\right),\\ &=\text{negative}.\end{aligned}$$

Now when $c_1(b-z)$ is big, then $e^{-c_1k(b-z)}$ is greater than $e^{-c_1h(b-z)}$, and consequently P_x has the sign of the first term, that is, positive. It follows that P_x changes from negative at the boundary to positive a short distance from the boundary and quickly becomes very small. Also it changes sign with z.

The positions of the maxima of S_y and their magnitudes can easily be calculated. For

$$S_y \propto e^{-kc_1d}-e^{-hc_1d}.$$

Fig. 5.



The condition for maxima is

$$-kc_1e^{-kc_1d}+hc_1e^{-hc_1d}=0,$$

whence

$$e^{-(h-k)c_1d}=\frac{k}{h},$$

i. e.,

$$c_1d=\frac{1}{h-k}\log_e\frac{h}{k}. \quad . \quad . \quad . \quad . \quad . \quad (160)$$

If we take $k=0{\cdot}375$, $h=\frac{7}{3}k$, the approximate values near $V^2=0{\cdot}86\,V_n^2$, we get

$$kc_1d=\frac{3}{4}\log_e\frac{7}{3}=0{\cdot}6354,$$

$$c_1d=\frac{0{\cdot}6354}{0{\cdot}375}=1{\cdot}955. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (161)$$

Thus it is not d we find but c_1d, and the value of d depends on the particular value of c_1 we choose near $c_1=\infty$. For example, from our graph we find that, at $V^2=0{\cdot}8\,V_n^2$,

$$\begin{aligned}c_1^2b^2&=12,\\ c_1b&=3{\cdot}46. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (162)\end{aligned}$$

Consequently,

$$\frac{d}{b}=\frac{1\cdot955}{3\cdot46}=0\cdot565. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (163)$$

This is not a big value of c_1b and therefore the maximum occurs nearer the centre than the boundary. For this moderate value of c_1b our approximations are not very good.

In order to examine the sharpness of the skin waves, it is necessary to solve equation (120) for values of V^2/V_n^2 near the critical value where $c_1=\infty$. For big values of c_1 this equation becomes

$$1-2e^{-2kR_1}+2e^{-2hR_1}=\frac{(1+k^2)^2}{4hk}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (164)$$

Now let the values of h and k when $c_1=\infty$ be denoted by h_1 and k_1, and let the corresponding value of V^2/V_n^2 be r. For large values of c_1 we next put

$$\frac{V^2}{V_n^2}=r-\epsilon, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (165)$$

and then

$$k^2=k_1^2+\epsilon, \qquad k=k_1+\frac{\epsilon}{2k_1}; \quad . \quad . \quad . \quad . \quad . \quad . \quad (166)$$

$$h^2=h_1^2+\frac{2}{7}\epsilon, \quad h=h_1+\frac{\epsilon}{7h_1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (167)$$

Consequently, the approximate equation connecting c_1b and ϵ is

$$2(e^{-2k_2R_1}-e^{-2h_1R_1})=\frac{4hk-(1+k^2)^2}{4hk}$$

$$=\frac{4h_1k_1-(1+k_1^2)^2+4\epsilon\left(\frac{h_1}{2k_1}+\frac{k_1}{7h_1}\right)-2\epsilon(1+k_1^2)}{4h_1k_1}.$$

Now, since h_1 is more than twice k_1, the term $e^{-2h_1R_1}$ is less than the square of $e^{-2k_1R_1}$ and is therefore negligible in comparison. Also

$$4h_1k_1=(1+k_1^2)^2.$$

Therefore

$$e^{-2k_1R_1}=\epsilon\left(\frac{1}{4k_1^2}+\frac{1}{14h_1^2}-\frac{4h_1k_1}{1+k_1^2}\right)$$

$$=0\cdot88\epsilon. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (168)$$

But we found that the position of maximum S_y is given by

$$e^{-(h-k)c_1d}=\frac{k}{h}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (169)$$

or, approximately, for $\sigma=0\cdot3$

$$e^{-\frac{4}{3}k_1c_1d}=0\cdot429. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (170)$$

Let us now find ϵ corresponding to the case where $d/b=0\cdot2$. Thus

$$2k_1R_1=10k_1c_1d=7\cdot5\left(\frac{4}{3}k_1c_1d\right),$$

and so

$$\epsilon=\frac{e^{-2k_1\mathrm{R}_1}}{0{\cdot}88}=\frac{(0{\cdot}429)^{7{\cdot}5}}{0{\cdot}88}$$

$$=0{\cdot}00202. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (171)$$

So the horns of the shear curve are at distances $\frac{1}{5}b$ from the free boundary when

$$\frac{\mathrm{V}^2}{\mathrm{V}_n{}^2}=r-0{\cdot}00202,$$

$$=0{\cdot}86-0{\cdot}00202, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (172)$$

assuming $0{\cdot}86$ to be the correct value of r.

It is evident then that the really sharp skin waves do not occur till V^2 has very nearly its critical value $r\mathrm{V}_n{}^2$, although the horned form of the shear curve appears for a relatively very much bigger value of ϵ.

It is interesting to find at what value of V^2 the horns begin to appear. When V^2 is small the shear curve is parabolic, with a distinct maximum at the middle of the rod. As V^2 increases there comes a state at which the central maximum flattens and begins to form a pair of horns, giving a minimum at the centre and two maxima on each side.

The condition for maxima or minima of S_y is

$$\frac{\partial \mathrm{S}_y}{\partial z}=0,$$

i. e.,

$$k\frac{\sinh c_1kz}{\cosh c_1kb}=h\frac{\sinh c_1hz}{\cosh c_1hb}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (173)$$

One root of this is, of course, $z=0$ for all values of V^2. But just at the point when the two horns are forming the equation must have a triple root at $z=0$. Now, expanding in powers of z, we get

$$\frac{k}{\cosh c_1kb}\{c_1kz+\tfrac{1}{6}c_1{}^3k^3z^3+\}=\frac{h}{\cosh c_1hb}\{c_1hz+\tfrac{1}{6}c_1{}^3h^3z^3+\}.$$

This gives a triple root provided that the first terms on the two sides are equal; that is,

$$\frac{k^2}{\cosh c_1kb}=\frac{h^2}{\cosh c_1hb}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (174)$$

By the use of the graph connecting R^2 and $\mathrm{V}^2/\mathrm{V}_n{}^2$, it can be shown that this is satisfied when

$$\left.\begin{aligned}\mathrm{V}^2&=0{\cdot}787\ \mathrm{V}_n{}^2,\\ \mathrm{R}_1{}^2&=10{\cdot}2,\\ \lambda/b&=2\pi/\mathrm{R}_1=2\text{ nearly.}\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (175)$$

Thus the horns on the S_y graph begin to appear when the wave-length is just about equal to the depth of the rod.

Another point of interest about these vibrations is that the total bending moment M on the section vanishes for a particular value of V^2. Equation (148 *a*) gives M, and when we use (117) to eliminate $\tanh c_1hb$ from M, we find that the equation simplifies to

$$\frac{\mathrm{M}}{2nb\mathrm{X}_1\mathrm{L}}=\frac{1}{h^2(1+k^2)}\frac{\mathrm{V}^4}{\mathrm{V}_n{}^4}\left\{\frac{2m}{m+n}\frac{\tanh k\mathrm{R}_1}{k\mathrm{R}_1}-\frac{1}{2}\frac{m-n}{m+n}(1+k^2)\right\}. \quad (176)$$

So M is zero when

$$\frac{\tanh kR_1}{kR_1}=\frac{m-n}{4m}(1+k^2)$$

$$=\frac{3}{20}(1+k^2). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (177)$$

This gives a fairly big value of R_1 and consequently there is little error in taking $\tanh kR_1=1$, and $k^2=0{\cdot}14$. Then

$$kR_1=\frac{20}{3(1+k^2)}=5{\cdot}85. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (178)$$

But near $V^2=0{\cdot}86\ V_n{}^2$ we have just found

$$\epsilon=1{\cdot}14e^{-2kR_1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (179)$$

So at the point where $M=0$

$$\epsilon=1{\cdot}14e^{-11{\cdot}7}<10^{-5}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (180)$$

It is clear then that M does not vanish till V^2 has very nearly reached the critical value at which the wave-length is infinitely small. The actual wave-length is

$$\lambda=\frac{2\pi}{c}=\frac{2\pi kb}{kR_1}=0{\cdot}400\,b, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (181)$$

which is one-fifth of the depth of the strip. These are just surface ripples.

The shear force F never changes sign and so, at this particular wave-length, there is a shear force without a bending moment.

In the region $V^2=0$ to $V^2=0{\cdot}86\ V_n{}^2$ the value of w is

$$w=LX_1\left\{\frac{1}{1+k^2}\cdot\frac{\cosh c_1kz}{\cosh c_1kb}-\frac{1}{2}\frac{\cosh c_1hz}{\cosh c_1hb}\right\}, \quad . \quad . \quad . \quad (182)$$

and when c_1b is big this becomes approximately

$$w=LX_1\left\{\frac{1}{1+k^2}e^{-c_1k(b\pm z)}-\frac{1}{2}e^{-c_1h(b\pm z)}\right\}, \quad . \quad . \quad . \quad . \quad (183)$$

the plus sign being taken where z is negative and the minus sign where z is postive.

At the surface, where $b\pm z=0$,

$$w=LX_1\left\{\frac{1}{1+k^2}-\frac{1}{2}\right\}$$

$$=LX_1\left\{\frac{1}{1{\cdot}14}-\frac{1}{2}\right\}\text{ nearly,}$$

$$=\frac{43}{114}LX_1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (184)$$

At a short distance from the boundary the dilatation term is negligible in comparison with the other term, and therefore

$$w=\frac{LX_1}{1{\cdot}14}e^{-c_1k(b\pm z)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (185)$$

Thus w keeps the same sign and decreases very quickly with the distance from the nearer boundary.

It is an interesting fact that the skin waves are predominantly distortion waves. The contribution of dilatation to w is 0·57 as much at distance 2·67 d from the boundary as the contribution of rotation at distance d; and, at the same distance from the boundary, where either contribution is small, they are not of the same order of magnitude.

Secondary Transverse Waves on a Rod.

The equation (113), which gives the relation between R^2 and V^2, has an infinite number of solutions for each value of V^2. So far all that we have got are the solutions which lie on one curve passing through the origin $R^2=0$, $V^2/V_n^2=0$. The complete solution is represented by an infinite number of curves of which we have so far found one branch, the most important branch because it includes vibrations of all wave-lengths.

To get the other solutions of (113) we use the form (125), with c for ic_1, and find

$$\frac{\sin(h-k)R}{\sin(h+k)R}=\frac{l-1}{l+1}=q \text{ say.} \qquad (186)$$

What we know about q is that it is zero when $V^2=0·86\,V_n^2$, it is positive and less than 1 when $0<V^2<0·86\,V_n^2$, and is negative and less than 1 when $V^2>0·86\,V_n^2$. One certain deduction we can make from (186) is therefore than $\sin(h-k)R$ is always less than 1 in magnitude and may be positive or negative according to the signs of q and $\sin(h+k)R$.

To build up a continuous series of solutions of (186), represented by one branch of the R^2 and V^2/V_n^2 curve, we can start with $V^2=0·86\,V_n^2$.

Then
$$l=1, \quad q=0,$$

$$\frac{\sin(h-k)R}{\sin(h+k)R}=0. \qquad (187)$$

For real values of R this has an infinity of solutions of the form

$$(h-k)R=s\pi, \qquad (188)$$

and then
$$(h+k)R=\frac{h+k}{h-k}s\pi. \qquad (189)$$

Every value of s from 1 to ∞ gives a branch of the curve connecting R^2 and V^2/V_n^2. The second branch of the curve—the first branch we have already got by another method—comes from taking $s=1$. Then

$$(h-k)R=(0·8685-0·3742)R$$
$$=0·4943R=\pi; \qquad (190)$$
$$(h+k)R=1·243R=2·514\pi; \qquad (191)$$
$$R=2·028\pi. \qquad (192)$$

Now we have shown that $\sin(h-k)\pi$ is never so great as 1 in magnitude as long as q is less than 1, and this is true from $V^2=V_n^2$ to $V^2=0$. Therefore, as V decreases to zero, $(h-k)\pi$ must lie between $\pi/2$ and $3\pi/2$. But, from the rough table of values of $(h+k)/(h-k)$ given below, it is clear that $(h+k)\mathrm{R}$ must increase up to infinity as V decreases to zero. This fact enables us to calculate easily a series of corresponding values of R for, every time that $(h+k)\mathrm{R}$ passes through a multiple of π, the angle $(h-k)\mathrm{R}$ must pass through π. Thus the smaller angle oscillates backward and forward through π while the larger, starting at $2{\cdot}514\pi$, passes successively through 3π, 4π, etc.

$\mathrm{V}^2/\mathrm{V}_n^2$	0	0·2	0·4	0·6	0·8	0·86	1
$(h+k)/(h-k)$	∞	24·0	10·6	5·5	3·08	2·514	1

The set of corresponding values are given by

$$\left.\begin{aligned}(h-k)\mathrm{R}&=\pi,\\(h+k)\mathrm{R}&=r\pi,\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (193)$$

where r is any integer from 3 to ∞. From these we find

$$\frac{h}{k}=\frac{r+1}{r-1},$$

whence

$$\frac{\mathrm{V}^2}{\mathrm{V}_n^2}=\frac{(r+1)^2-(r-1)^2}{(r+1)^2-\dfrac{n}{m+n}(r-1)^2}=\frac{(r+1)^2-(r-1)^2}{(r+1)^2-\frac{2}{7}(r-1)^2}. \quad . \quad (194)$$

Also

$$(h^2-k^2)\mathrm{R}^2=r\pi^2,$$

from which

$$\mathrm{R}^2=\frac{m+n}{m}r\frac{\mathrm{V}_n^2}{\mathrm{V}^2}\pi^2=\frac{7}{5}r\frac{\mathrm{V}_n^2}{\mathrm{V}^2}\pi^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad (195)$$

The following table is calculated from these :—

TABLE C 1.

r	3	4	5	6	7	8	15	∞
$\mathrm{V}^2/\mathrm{V}_n^2$ . . .	0·808	0·713	0·636	0·573	0·521	0·478	0·300	0
R^2/π^2	5·20	7·85	11·0	14·6	18·8	23·4	70	∞
$2k\mathrm{R}/\pi$ . . .	2	3	4	5	6	7	14	∞
$2h\mathrm{R}/\pi$ . . .	4	5	6	7	8	9	16	∞
R^2	51·3	77·5	108·6	144·2	185·6	231	691	∞

In continuing this table beyond $V^2=0{\cdot}86V_n^2$, account must be taken of the fact that $\sin(h-k)R$ is zero at this critical point but $\sin(h+k)R$ is not zero. Nevertheless, this does not interrupt the sequence and we can complete the table by going down to $r=1$. Here is the completed table.

TABLE C 2.

r	1	2
V^2/V_n^2	1	0·918
R^2/π^2	1·40	3·05
$2kR/\pi$	0	1
$2hR/\pi$	2	3
R^2	13·82	30·1

This branch of the curve goes on beyond $V=V_n$, but in this region k is imaginary. It is necessary now to return to equation (113) and put ik_1 for k. The new equation is

$$\frac{\tanh k_1R}{\tan hR}=-\frac{(1-k_1^2)^2}{4hk_1}. \quad \ldots\ldots\ldots \quad (196)$$

This makes $\tan hR$ negative right up to $V=V_D$ and, since hR is π at $V=V_n$, continuity requires that

$$\pi \geqq hR \geqq \frac{\pi}{2}. \quad \ldots\ldots\ldots\ldots \quad (197)$$

When $V^2=2V_n^2$, then

$$\left.\begin{aligned} 1-k_1^2=2-\frac{V^2}{V_n^2}=0,\\ h^2=\frac{3}{7}. \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (198)$$

So hR reaches the value $\pi/2$ at this point. Again when $V=V_D$ then $h=0$ and consequently, since $\tanh k_1R$ is not infinite,

$$\tan hR=0. \quad \ldots\ldots\ldots\ldots \quad (199)$$

The only conclusion from this is that

$$\left.\begin{aligned} hR=\pi,\\ R=\infty. \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (200)$$

and

So, for this region, we find the values in the following table :—

TABLE C 3.

V^2/V_n^2 ...	1	1·5	2	2·5	3	3·5
R^2/π^2	1·40	0·505	0·582	0·974	2·86	∞
R^2	13·82	4·99	5·76	9·61	28·1	∞

We have now got a sufficiently complete set of values of R^2 and $V^2/V_n{}^2$ to plot the second branch of the curve in the region where V^2 is positive. Other branches can be got by taking $s=2$, 3, etc. in (188) and (189) and using the same methods. The first of these equations shows that an increase of 1 in s makes an increase of $\pi/(h-k)$ in R, *i. e.*, $2{\cdot}028\,\pi$ by (192). Further, reasoning similar to that already used will show that hR becomes 2π at $V^2=V_D{}^2$, and therefore R^2 becomes infinite again. It is possible to guess then, without elaborate arithmetic, what the forms of the next two or three branches are like.

On all these branches we have just been considering, that is, the second, third, etc., since R^2 is positive it follows that c is real, and so the displacements and stresses all have a factor $e^{c(x-Vt)}$ or all a factor $e^{-c(x-Vt)}$. Also, even on the second branch, the least value of R^2 is about 5 and consequently the least R is about 2·23. The exponential factor for this R is

$$e^{\pm 2{\cdot}23\frac{x-Vt}{b}}.$$

In a distance $2b$ along the x-axis this increases or decreases in the ratio $e^{4{\cdot}46}$ ($=86{\cdot}5$) to one. Stresses with such exponential factors then are very concentrated stresses, similar to the stresses in skin waves, with this difference however, that they can travel along the rod and thus occur at any section of the rod.

The distance, measured along the rod, from a point where the stress is measurable to a point where the elastic limit is reached, is not more than three times the depth of the rod for any ordinary material, even for the smallest value of c.

Waves with the factor $e^{c(x-Vt)}$ have no beginning and no end, and consequently it is theoretically impossible to generate such a wave in a rod. But in practice we could assume that the wave does not exist in front of the point of the rod where the stresses are so small as to be completely negligible, the word front here being used for the region where the exponential function fades away. There is nothing unreasonable then in assuming that these exponential waves have a beginning, and, with this assumption, such waves could be generated by proper forces applied at one end of a rod which was initially at rest. If rupture is not to occur the exponential form of wave can be generated for only a very short time, the time taken by the wave to travel a few times the depth of the rod. This would only be of the nature of a sudden blow. When the end forces cease entirely then we are confronted with the problem of a free rod with given displacements and velocities, and the subsequent states can be determined by methods not yet worked out on the present theory. This is a problem for further investigation.

When a tuning-fork is struck it is obvious that one or more exponential waves are at work, and when the blow is over the free vibrations begin.

To complete the investigation of waves on a thin rod we have to examine what other branches of the curve connecting R^2 and V^2 exist when $V^2=-U^2$.

The equation giving the relation between these variables is

$$\frac{\tan k\mathrm{R}}{\tan h\mathrm{R}}=\frac{(1+k^2)^2}{4hk}=\mathrm{N}\text{ say}, \quad . \quad . \quad . \quad . \quad . \quad . \qquad (201)$$

where now

$$k^2=1+\frac{\mathrm{U}^2}{\mathrm{V}_n{}^2}, \quad h^2=1+\frac{2}{7}\frac{\mathrm{U}^2}{\mathrm{V}_n{}^2}(\sigma=0{\cdot}3), \quad . \quad . \quad . \quad . \qquad (202)$$

and

$$\begin{aligned}\frac{k^2}{h^2}&=\left(1+\frac{\mathrm{U}^2}{\mathrm{V}_n{}^2}\right)\bigg/\left(1+\frac{2}{7}\frac{\mathrm{U}^2}{\mathrm{V}_n{}^2}\right)\\ &\left.\begin{aligned}&=1\text{ when }\mathrm{U}=0,\\ &=\frac{7}{2}\text{ when }\mathrm{U}=\infty.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (203)\end{aligned}$$

Clearly k/h varies from 1 to 1·871 as U varies from 0 to ∞. Also N varies from 1 to ∞ in the same range.

We can get all we need out of (201) by a method we used earlier. Thus

$$\frac{\sin (k-h)\mathrm{R}}{\sin (k+h)\mathrm{R}}=\frac{\mathrm{N}-1}{\mathrm{N}+1}\overset{=}{<}1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (204)$$

This last equation shows that $\sin (k-h)\mathrm{R}$ can never have the magnitude unity except, possibly, at $\mathrm{U}=0$ or $\mathrm{U}=\infty$. This means that the angle $(k-h)\mathrm{R}$ cannot vary by more than π. So let

$$(k-h)\mathrm{R}=s\pi+\alpha,$$

where α can vary between the limits $\pm\pi/2$.

If s is not zero then the least value of $(k-h)\mathrm{R}$ is a finite angle, and therefore

$$(k+h)\mathrm{R}=\frac{k+h}{k-h}(k-h)\mathrm{R}$$

is an angle which varies from ∞ at $\mathrm{U}=0$ to $2{\cdot}871(k-h)\mathrm{R}/0{\cdot}871$ at $\mathrm{U}=\infty$. Thus the angle $(k+h)\mathrm{R}$ passes through an infinite range and, in so doing, passes through multiples of π. But when $(k+h)\mathrm{R}$ passes through a multiple of π, equation (204) shows that $(k-h)\mathrm{R}$ also passes through a multiple of π, and the only possible multiple for the latter angle is $s\pi$. Thus the smaller angle passes backwards and forwards through $s\pi$ while the larger ranges from ∞ to something in the neighbourhood of $2{\cdot}87s\pi/0{\cdot}871$.

We get a series of corresponding values of R^2 and U^2 by taking

$$\left.\begin{aligned}(k-h)\mathrm{R}&=s\pi,\\ (k+h)\mathrm{R}&=r\pi,\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (205)$$

whence

$$\left.\begin{aligned}k\mathrm{R}&=\tfrac{1}{2}(r+s)\pi,\\ h\mathrm{R}&=\tfrac{1}{2}(r-s)\pi,\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (206)$$

$$(k^2-h^2)\mathrm{R}^2=rs\pi^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (207)$$

i. e.,

$$\frac{5}{7}\frac{\mathrm{U}^2}{\mathrm{V}_n{}^2}\mathrm{R}^2=rs\pi^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (208)$$

Also
$$\frac{U^2}{V_n^2}=\frac{28rs}{7(r-s)^2-2(r+s)^2}, \quad . \quad . \quad . \quad . \quad . \quad (209)$$

and therefore
$$\frac{20}{\pi^2}R^2=7(r-s)^2-2(r+s)^2. \quad . \quad . \quad . \quad . \quad . \quad (210)$$

Each integral value of s, from 1 upwards, gives a different branch of the curve. On any one branch r has all the possible integral values that correspond to positive values of R^2. The smallest value of r that makes R^2 positive determines in what quadrants the angles kR and hR arrive as U approaches ∞. Since $N=\infty$ when $U=\infty$ we deduce from (20) that, at this stage, either kR is an odd multiple of $\pi/2$ or else hR is a multiple of π and R itself is zero. It will be found, however, by actual trial, and it can be proved from our equations, that $R=0$ only when kR is an odd multiple of $\pi/2$, and never when hR is a multiple of π.

Here are tables of values of the essential variables for the second and third branches.

TABLE D.

Second branch for $V^2=-U^2$; $s=1$.

r	—	4	5	6	7	8	9	10	15	20
$20R^2/\pi^2$ ..	0	13	40	77	124	181	248	325	860	1645
U^2/V_n^2...	∞	8·615	3·500	2·182	1·581	1·237	1·016	0·8615	0·4884	0·3404
$2kR$	5π	5π	6π	7π	8π	9π	10π	11π	16π	21π
$2hR$	$2{\cdot}672\pi$	3π	4π	5π	6π	7π	8π	9π	14π	19π

TABLE E.

Third branch for $V^2=-U^2$; $s=2$.

r	—	7	8	9	10	15	20	30
$20R^2/\pi^2$.	0	13	52	101	160	605	1300	3440
U^2/V_n^2..	∞	30·15	8·615	4·990	3·500	1·388	0·8615	0·488
$2kR$	9π	9π	10π	11π	12π	17π	22π	32π
$2hR$	$4{\cdot}81\pi$	5π	6π	7π	8π	13π	18π	28π

All the branches of the curve connecting R^2 and V^2/V_n^2 that we have so far dealt with are shown in fig. 6. The value of R^2 for the first branch in the region, where V^2 is negative, is so small that the curve almost runs along the axis of V^2/V_n^2.

Several particular solutions of the vibration problem have their own special interest, and a very interesting solution is the one where $V=V_n$ on the second branch (Table C3). In this case

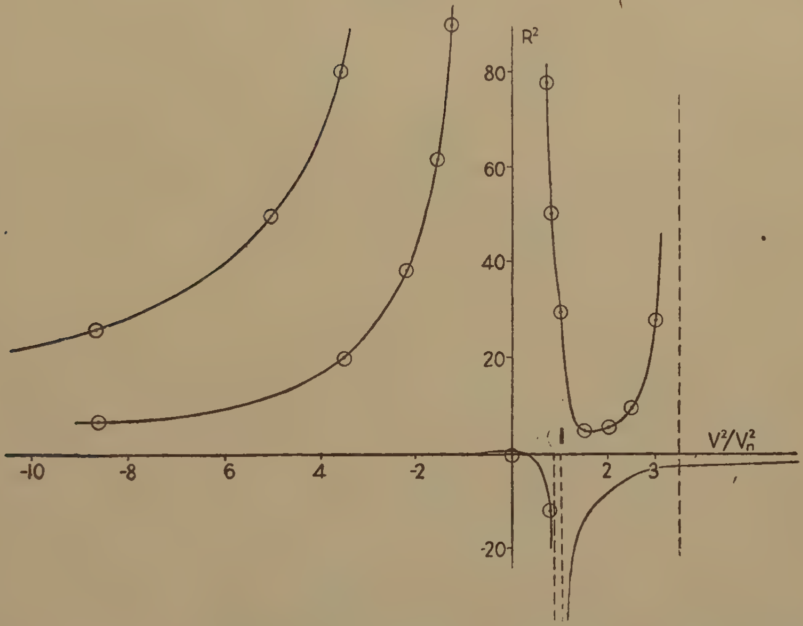
$$k=0, \quad chb=\pi, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (211)$$

and therefore

$$\left.\begin{aligned} u&=\mathrm{T}e^{c(x-\mathrm{V}_nt)}\sin\frac{\pi z}{b},\\ w&=h\mathrm{T}e^{c(x-\mathrm{V}_nt)}\left(2+\cos\frac{\pi z}{b}\right),\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (212)$$

$$\left.\begin{aligned} \mathrm{P}_x&=\mathrm{P}_0e^{c(x-\mathrm{V}_nt)}\sin\frac{\pi z}{b},\\ \mathrm{S}_y&=\mathrm{B}e^{c(x-\mathrm{V}_nt)}\left(1+\cos\frac{\pi z}{b}\right),\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (213)$$

Fig. 6.



where P_0 and B are constants and proportional to T. The only term in the displacements that gives rise to a distortion is the term

$$2hTe^{c(x-\mathrm{V}_nt)} \text{ in } w,$$

and this is a purely plane wave, from which only the partial stress

$$\mathrm{S}_y=\mathrm{B}e^{c(x-\mathrm{V}_nt)}$$

results.

This particular transverse motion may therefore be regarded as merely a plane distortion wave accompanied by just the necessary dilatation wave to satisfy the boundary conditions.

Rod with a Circular Section.

Owing to the length of this paper we shall do no more than quote results for a cylindrical rod.

In some respects the cylindrical rod is a better rod to deal with than the flat strip, but in others it is worse. Bessel functions now appear instead of circular functions, but one advantage is that several different types of oscillation appear under what is practically one form.

With co-ordinates x along the rod and r, θ in the section, the following displacements are possible :—

for dilatation [J_n denotes the Bessel function of order n]

$$u=2HJ_n(chr)\cos n\theta, \quad \ldots \ldots \quad (214)$$

$$s=hH\{J_{n-1}(chr)-J_{n+1}(chr)\}\cos n\theta, \quad \ldots \quad (215)$$

$$\eta=hH\{J_{n-1}(chr)+J_{n+1}(chr)\}\sin n\theta\,; \quad \ldots \quad (216)$$

for rotation

$$u=k(K_1-K_2)J_n(ckr)\cos n\theta, \quad \ldots \ldots \quad (217)$$

$$s=\{K_1J_{n-1}(ckr)+K_2J_{n+1}(ckr)\}\cos n\theta, \quad \ldots \quad (218)$$

$$\eta=\{K_1J_{n-1}(ckr)-K_2J_{n+1}(ckr)\}\sin n\theta. \quad \ldots \quad (219)$$

The total displacements are the sums of the dilatation and rotation terms, and stresses can be derived from these which will satisfy the boundary conditions over the cylindrical surface.

By taking different values of n we get different types of vibrations. Thus

$n=0$ gives longitudinal vibrations,

$n=1$ gives transverse vibrations.

Any other value of n gives vibrations with n nodal diameters, types of vibration that have not such simple physical characteristics as the first two.

When $n=1$, that is, for transverse vibrations, we find, for small values of V^2,

$$c^2K^2=-\frac{V^2}{V_E{}^2},$$

where $K^2=r^2/4$ is the square of the radius of gyration of the section about the neutral axis.

Also, in every case, that is, for all values of n, the vibrations develop into skin waves when

$$(1+k^2)^2=4hk,$$

exactly as for the flat strip. These skins waves travel along the surface in the direction parallel to the x-axis.

There can be no doubt that the curve connecting R^2 and $V^2/V_n{}^2$ would be almost exactly the same as for the flat strip, the main difference being that, for small values of $V^2/V_n{}^2$, E now takes the place of $E/(1-\sigma^2)$.

It is well worthy of special notice that even longitudinal waves develop into skin waves when the wave-length is infinitely small, and the velocity of these is $\sqrt{0{\cdot}86}\,V_n\simeq 0{\cdot}928\,V_n$, which is very different from V_E of the accepted theory. As λ varies from 0 to ∞ the velocity V varies from the above skin-wave velocity to V_E.

We have now accumulated enough material to test the accuracy of the differential equations obtained by a combination of exact dynamical theory and guesses at the stress distribution across the section of the rod.

The differential equation (41) gives nothing but the straight line relation

$$K^2c^2 = -\frac{V^2}{V_E^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (220)$$

between K^2c^2 and V^2, on the understanding that the wave factor is $e^{c(x-Vt)}$ and not $e^{ic(x-Vt)}$. This equation corresponds to the tangent at the origin to the first branch of our exact curve connecting R^2 and V^2/V_n^2. A glance at figs. 3 and 4 will show how far wrong this is except for small values of V^2.

The differential equation containing the usual rotary inertia term gives

$$K^2c^2 = -\frac{V^2}{V_E^2 - V^2} . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (221)$$

This has the same tangent at the origin and has two asymptotes, a vertical one at $V^2 = V_E^2$ and a horizontal asymptote

$$K^2c^2 = 1$$

at $V^2 = -\infty$.

This graph looks a little more like the graph of the first branch of the exact curve, but is still very far from the truth for all but small velocities.

The differential equation (86) gives

$$K^2c^2 = -\frac{V_n^2 V^2}{(V_f^2 - V^2)(V_n^2 - V^2)} . \quad . \quad . \quad . \quad . \quad . \quad (222)$$

As it stands this equation gives a graph of very much the same form as the first branch of the exact curve, but it fails by giving an asymptote at $V^2 = V_n^2$ instead of at $V^2 = 0{\cdot}86\,V_n^2$. This infinite value of K^2c^2 at the velocity of skin waves could not have been predicted by the method of guessing at the stress distribution because there is nothing in the Bernoulli-Euler approach to the solution which suggests skin waves or any stress distribution of that nature.

By a slight change in the form of (222) this equation will give a graph with all the characteristics of the first branch of the exact curve. In fact, the two graphs are almost indistinguishable; both have a maximum value of K^2c^2 of nearly the same magnitude at nearly the same value of U^2.

The changed equation is, for a real thin rod,

$$K^2c^2 = -\frac{V_D^2}{V_E^2} \frac{V_n^2 V^2}{(V_n^2 - V^2/0{\cdot}86)(V_D^2 - V^2)} , \quad . \quad . \quad . \quad (223)$$

but for comparison with the flat strip it is necessary to use $V_{E'}^2$, instead of V_E^2, that is, $E/(1-\sigma^2)$ for E.

There is little error in taking 0·86=6/7 for arithmetical work. Then the equation suitable for the flat strip is, with $\sigma=0{\cdot}3$ and $K^2=\frac{1}{3}b^2$,

$$R^2=-\frac{3V^2}{V_{E'}^2}\frac{1}{\left(1-\frac{7}{6}\frac{V^2}{V_n^2}\right)\left(1-\frac{2}{7}\frac{V^2}{V_n^2}\right)}$$

$$=-\frac{3n(1-\sigma^2)}{E}\frac{V^2}{V_n^2}\cdot\frac{1}{\left(1-\frac{7}{6}\frac{V^2}{V_n^2}\right)\left(1-\frac{2}{7}\frac{V^2}{V_n^2}\right)}$$

$$=-\frac{21}{20}\frac{V^2}{V_n^2}\frac{1}{\left(1-\frac{7}{6}\frac{V^2}{V_n^2}\right)\left(1-\frac{2}{7}\frac{V^2}{V_n^2}\right)} \quad . \quad . \quad . \quad . \quad . \quad (224)$$

Corresponding values of R^2 and V^2/V_n^2 are here tabulated, together with the values of R^2 from tables A and B.

V^2 positive.

V^2/V_n^2	0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·86
R^2 (Eqn. (224))	0	0·122	0·290	0·530	0·889	1·47	2·53	5·01	16·3	∞
R^2 (exact)	0	0·120	0·288	0·510	0·85	1·38	2·38	4·55	12·0	∞

V^2 negative.

$-V^2/V_n^2$	0·5	1	2	3	4	7	∞
R^2 (Eqn. (224)).	0·290	0·377	0·401	0·377	0·346	0·267	0
R^2 (exact)	0·281	0·365	0·371	0·341	0·300	0·222	0

When one considers the complexity of the equation that gives the exact values of R^2 and the simplicity of (224), the agreement shown in the above table is astonishing for so large a range of values of V^2. The proportional error in R is, of course, only about half of that in R^2, and this error in R becomes serious only in the neighbourhood of the skin waves. It is never very big for negative values of V^2. At $V^2=-\infty$ we see from our previous results that

$$\frac{R^2V^2}{V_n^2}\to-\frac{\pi^2}{4}=-2{\cdot}47 \text{ (from exact equations)} \quad . \quad . \quad . \quad (225)$$

and

$$\frac{R^2V^2}{V_n^2}\to-\frac{63}{20}=-3{\cdot}15 \quad \text{(Eqn. (224))}. \quad . \quad . \quad . \quad . \quad . \quad (226)$$

Even here the difference is not very great.

It is clear then that the first branch of the exact curve connecting R^2 and V^2 is fairly accurately represented by equation (224), and if it be granted that this first branch includes all the primary free vibrations it follows that the differential equation (86), from which (222) is derived, can easily be modified so that it represents fairly precisely the free vibrations of a rod. This modified differential equation could then replace (41) or (55) when more precision is required than either of these equations can give.

The suggested modified equation is, for a thin rod,

$$K^2\left\{V_D^2\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial t^2}\right\}\left\{\frac{7}{6V_n^2}\frac{\partial^2 W}{\partial t^2}-\frac{\partial^2 W}{\partial x^2}\right\}=\frac{V_D^2}{V_E^2}\frac{\partial^2 W}{\partial t^2}. \quad . \quad (227)$$

This equation corresponds to (223).

Equation (223) shows that (227) has solutions for W with a factor $e^{c(x-Vt)}$ for values of V greater than $0{\cdot}86\,V_n^2$; but, although the general character of the graph giving K^2c^2 against V^2/V_n^2 is not unlike the second branch of the exact curve in this region, it fails by having an asymptote at $V^2=0{\cdot}86\,V^2$ instead of at $V^2=0$. Yet the second asymptote is at the right place at $V^2=V_D^2$. Nevertheless, no claim is here made that (227) can be used for any but the primary oscillatory motions of the rod, *i. e.*, those belonging to the first branch of the characteristic curve.

No use could be made of either equation (86) or (227) without expressions for the shear force F and the bending moment M.

The expressions for these that are consistent with the equation (86) are

$$F=\int\rho A\frac{\partial^2 W}{\partial t^2}dx \quad \text{by (61)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (228)$$

and

$$\begin{aligned} M&=\iint P_x z\,dA \\ &=\iint fz\frac{\partial u}{\partial x}dA \quad \text{by (79)} \\ &=fAK^2\left\{\frac{1}{V_n^2}\frac{\partial^2 W}{\partial t^2}-\frac{\partial^2 W}{\partial x^2}\right\}. \quad . \quad . \quad . \quad . \quad . \quad (229) \end{aligned}$$

by (81 and (85).

The uncertain factor f gives a corresponding uncertainty to M. Since it is probably best to have the equation right for the usual long waves the constant E could be taken for f.

For the modified, and improved, equation (227) the expression for F remains unaltered; but

$$M=EAK^2\left\{\frac{7}{6}\frac{1}{V_n^2}\frac{\partial^2 W}{\partial t^2}-\frac{\partial^2 W}{\partial x^2}\right\}. \quad . \quad . \quad . \quad . \quad (230)$$

This makes M proportional to μ' and gives the correct constant factor for long waves. It is, of course, impossible to make this approximate equation give any idea of the skin stresses.

Solutions of the Exact Equations with $V>V_D$.

The equation (224), derived from the approximate differential equation, gives a real value of R^2 for every real value of V^2, even for $V^2>V_D^2$. This suggests the possibility that the exact equation (113) has solutions for these large values of V^2. We shall now show that there are such solutions and estimate what physical meaning they have.

In equation (113) we assume that $V^2>V_D{}^2$ and make the substitutions

$$k=ik_1,\quad h=ih_1,\quad c=ic_1. \qquad (231)$$

The equation then becomes

$$\frac{\tan k_1R_1}{\tan h_1R_1}=-\frac{(k_1{}^2-1)^2}{4h_1k_1}. \qquad (232)$$

As V varies from V_D to ∞ the number h_1 varies from 0 to ∞, while k_1 varies from $\sqrt{2\cdot 5}$ to ∞. At the start, when $V=V_D$,

$$\frac{\tan k_1R_1}{k_1R_1}=-\frac{(k_1{}^2-1)^2}{4k_1{}^2}\frac{\tan h_1R_1}{h_1R_1}$$

$$=-\frac{1\cdot 5^2}{4\times 2\cdot 5}=-0\cdot 225. \qquad (233)$$

This has an infinity of solutions which lie on different branches of the $(R^2)(V^2/V_n{}^2)$ curve. The first solution is

$$k_1R_1=0\cdot 830\pi=2\cdot 61. \qquad (234)$$

while, of course,

$$h_1R_1=0. \qquad (235)$$

As V increases the angle h_1R_1 must move into the first quadrant, because $\tan h_1R_1$ and $\tan k_1R_1$ have a negative ratio. Further, it can be shown that, on this one branch of the curve, neither of the angles h_1R_1 or k_1R_1 moves out of its original quadrant, and only at $V=\infty$ does k_1R_1 reach the value $\pi/2$.

So in the range where V varies from V_D to ∞

k_1R_1 varies from $0\cdot 830\pi$ to $0\cdot 5\pi$;
h_1R_1 varies from 0 to $0\cdot 267\pi$;
$R_1{}^2$ varies from $0\cdot 276\pi^2$ to 0;
$R_1{}^2\times V^2/V_n{}^2$ varies from $0\cdot 965\,\pi^2$ to $0\cdot 25\,\pi^2$;
and $\lambda/2b$ varies from $1\cdot 905$ to ∞.

Thus we find that there are mathematical solutions of our equations when $V>V_D$. But what do these mean physically? Can these vibrations be generated by local actions on a thin rod initially at rest? I think the answer is no; but the reasons will be more apparent for this answer after we have dealt with the question of normal functions for a thin rod.

This V that occurs in our equations is not necessarily the velocity of an elastic impulse or of energy. It is merely a wave-form velocity. If ϕ and ψ denote respectively dilatation and rotation potentials we can be sure that the *stuffs* represented by ϕ and ψ carry energy with them. In the case of the true exponential waves, that is, waves with the factor $e^{c(x-Vt)}$ with c and V real, there is no doubt that the energy flows with the wave form and at the same velocity V. But in the case of sine waves, with the factor $e^{ic(x-Vt)}$, each wave-length contains the same energy throughout the motion, and such flow of energy as there is could be merely a forward and backward surging along the rod, and its velocity need not be V. By our present method it is not at all easy to identify any element $\delta\phi$ or $\delta\psi$ and say where it has arrived at any instant. But we can certainly assert that any solution of the differential equations that implies an energy flow at a velocity greater than V_D is physically impossible.

Since these waves with $V > V_D$ satisfy the differential equations and also the boundary conditions at the surface (these last conditions being the conditions that no energy passes into the rod from the sides), it must follow that the actual energy flow in these waves is physically possible, whence we conclude that V is not the rate of flow of energy for one of these waves.

It is worth while verifying this last statement in one easy case, one of the cases where V is nearly ∞.

We assume that the rotation function is

$$\psi = K \cos qz \sin cx \sin pt, \quad . \quad . \quad . \quad . \quad . \quad . \quad (236)$$

and that c is infinitesimal while p, which equals cV, is finite. We assume that stresses with a factor c are negligible.

This rotation function gives displacements

$$u = \frac{\partial\psi}{\partial x} = -qK \sin qz \sin cx \sin pt, \quad . \quad . \quad . \quad . \quad (237)$$

$$w = -\frac{\partial\psi}{\partial x} = 0 \,; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (238)$$

and stresses

$$S_n = n\left(\frac{\partial^2\psi}{\partial z^2} - \frac{\partial^2\psi}{\partial x^2}\right) = -nq^2\psi. \quad . \quad . \quad . \quad . \quad . \quad (239)$$

$$P_x = P_z = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (240)$$

The boundary conditions

$$P_z = 0, \quad S_y = 0, \quad \text{at } z = b \quad . \quad . \quad . \quad . \quad . \quad . \quad (241)$$

are satisfied, without the use of a dilatation function, if

$$\cos qb = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (242)$$

one solution of which is

$$qb = \frac{\pi}{2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (243)$$

The differential equation for ψ now gives

$$-q^2 - c^2 = -\frac{p^2}{V_n^{\,2}}$$

or

$$q^2 = \frac{p^2}{V_n^{\,2}} \text{ nearly.} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (244)$$

So far the only assumption about c is that it is small, and clearly all the equations are satisfied whether it is real or imaginary. Now

$$V = p/c$$

and a real p and real c gives a real V ; whereas a real p and imaginary c gives an imaginary V, and thus brings the wave on the first branch of the $(R^2)(V^2/V_n^2)$ curve where $U=\infty$. At the same time $K \sin cx$ is changed into $K_1 \sinh cx$.

The strain energy in a length δx of the flat strip is, per unit width,

$$\tfrac{1}{2}\delta x \int_{-b}^{b} \frac{S_y^2}{n}\, dz = nbq^4K^2 \sin^2 cx \sin^2 pt \, . \, \delta x, \quad . \quad . \quad . \quad (245)$$

and the kinetic energy in the same element is

$$\tfrac{1}{2}\rho\delta x \int_{-b}^{b} \left(\frac{\partial u}{\partial t}\right)^2 dz = p^2\rho bq^2K^2 \sin^2 cx \cos^2 pt \, . \, dx,$$

which $$= nbq^4K^2 \sin^2 cx \cos^2 pt \, . \, \delta x, \quad . \quad . \quad . \quad . \quad (246)$$

since $$q^2 = \frac{p^2}{V_n^2} = \frac{\rho p^2}{n}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (247)$$

So the total energy in δx is

$$nbq^4K^2 \delta x \sin^2 cx, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (248)$$

which is independent of t to the approximations used.

This shows that there is practically no flow of energy along the rod when V is nearly ∞.

Observe that, if $c=0$ exactly, then

$$\psi = K \cos qz \sin pt, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (249)$$

$$S_y = -nq^2K \cos qz \sin pt \text{ exactly,}$$

$$F = \text{shear force per unit width,}$$

$$= -2nqK \sin pt,$$

$$q = \pi/2b = p/V_n. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (250)$$

This represents a shear force which is the same at every point of the rod but which has the definite period

$$\tau = 2\pi/p = 2\pi b/V_n. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (251)$$

These are very rapid oscillations for a rod of any reasonable dimensions.

Thus a periodic shear force can exist in a rod without any transverse motion. The only motion is a longitudinal forward and backward motion of the half of the rod on one side of the neutral axis and the contrary motion of the other half. The energy in each element δx of the rod remains constant and merely changes from strain energy into kinetic energy and back again into strain energy as time goes on. Thus, for $V=\infty$, there is no flow of energy along the rod during the vibration, and there is very little flow when V is big but not infinite.

The same conclusion concerning the flow of energy could be reached if, instead of the standing wave in (236), the progressive wave

$$\psi = K \cos qz \cos (cx - pt)$$

were used. We must conclude then that both the standing wave and progressive wave are physically possible modes of vibration of the rod when it is once started. The only question is : in what manner can they be started? We shall try to answer this after dealing with the normal modes.

The branch of the $(R^2)(V^2/V_n^2)$ curve that we have just been dealing with continues into the region where $V<V_D$. In this region h is real and therefore ih replaces h_1 in our equations, and this turns $\tan h_1R_1/h_1R_1$ into $\tanh hR_1/hR_1$. Without going into details of the calculations we tabulate below the essential quantities for this branch.

TABLE F. $V_n^2<V^2<\infty$.

V^2	∞	V_D^2	$2V_n^2$	V_n^2
k_1R_1	$\pi/2$	$0{\cdot}830\pi$	π	$\pi/2$
R_1^2	0	$0{\cdot}276\pi^2$	π^2	∞
$R_1^2V^2/V_n^2$	$\pi^2/4$	$0{\cdot}965\pi^2$	$2\pi^2$	∞

Thus the line $V^2=V_n^2$ is an asymptote to the branch of the characteristic curve represented in the above table, as it is to all branches that go to $V=\infty$.

The vibrations and the state of stress near $V^2=V_n^2$ on the above branch is interesting from a physical point of view. We state briefly what they are here.

We are assuming that V^2 is just greater than V_n^2 and consequently that c_1 is very big, while c_1k_1b is nearly $\pi/2$. Then it can be shown that

$$u=0,\quad c_1w=G\cos c_1(x-Vt)\cos\frac{\pi z}{2b},$$

$$S_y=nG\sin c_1(x-Vt)\cos\frac{\pi z}{2b},$$

$$P_x=0,\quad P_z=0,$$

are good approximations to the truth. Thus the only stress is a very rapidly alternating shear stress of very short wave-length, the distribution of this stress across the section being not unlike the parabolic distribution for very long waves. The displacement is practically all transverse, and even that is small. Consequently these vibrations may be regarded as transverse shivers.

The next curve of this series has $k_1R_1=3\pi/2$, 2π, $3\pi/2$, at $V^2=\infty$, $2V_n^2$, V_n^2 respectively.

Normal Modes of Vibration for a Given Rod.

The normal modes of free vibration of an elastic body are those modes which are possible for the given boundary conditions, and in which the displacements and stresses have the same factor $\sin(pt+\alpha)$. The simple functions that we have found for the flat strip, while satisfying the boundary conditions on the broad faces, are not capable of satisfying the end conditions as well. But, if we are content, as in the ordinary theory based on the equation (41), to replace the very exacting end conditions by simpler ones involving only W, $\partial W/\partial x$, F and M, then normal functions can be found by means of the characteristic $R^2(V^2/V_n^2)$ curves (which we shall in future call the characteristic curves) for the strip or rod.

The normal functions on this exact theory have not the beautifully clean precision of those derived from equation (41). This precision is, of course, due to the fact that (41) has smoothed out the complexities of the real physical problem of a vibrating rod and replaced the more exact equations by one which has no serious mathematical difficulties. The real problem has mathematical difficulties which may possibly deter any but a very assiduous person from solving exactly any problem of a vibrating rod. Nevertheless it is worth while finding out the magnitude of the errors in the well-established theory.

It will shorten the argument to have names for the different branches of the characteristic curves. The two parts of the first branch (the one that passes through the origin), where V^2 is positive and where V^2 is negative, we shall call the curves α and β respectively. The branches which extend beyond $V=V_D$ and are asymptotic to the line $V^2=V_n^2$ we shall call the curves ω_1, ω_2, etc., ω_1 being the branch nearest the V^2/V_π^2 axis.

By means of the characteristic curves it is possible to express W, F, M, as functions of V, a factor $He^{c(x-Vt)}$, or the corresponding real factor when $c=ic_1$, being always understood.

If p is given then $R^2\times V^2/V_n^2$ is given for the rod. So all possible vibrations with a given p are to be found at the intersections of the hyperbola

$$\frac{R^2V^2}{V_n^2}=\frac{c^2b^2V^2}{V_n^2}=-\frac{b^2}{V_n^2}p^2$$

with the characteristic curves.

If $-R^2V^2<\pi^2V_n^2/4$, the above hyperbola meets each of the curves α and β once. (It is easy to verify, by taking numerical values from the curve β, that R^2U^2/V_n^2 decreases from $U=0$ to $U=\infty$, where $R^2U^2/V_n^2=\pi^2/4$.) Each point of intersection corresponds to two values of c, namely $\pm c$, and consequently there are four possible vibrations with a given p within these limits. Corresponding to the value of p the complete expression for W has the form

$$W=\{L_1f_1\cos c_1x+L_2f_2\sin c_1x+L_3f_3\cosh c_2x+L_4f_4\sinh c_2x\}\sin(pt+\gamma),$$

where the L's are constants and the f's are functions of the c's. The quantities F and M can be expressed in the same form with the same values of the L's but with different values of the f's.

Modes of vibration symmetrical about the mid-point of the rod, taken as origin for x, involve only the even functions $\cos c_1x$ and $\cosh c_2x$ in W and M and the odd functions in F.

In any case, the end conditions, two for each end, give four equations connecting the L's and the constants $c_1 l$ and $c_2 l$, where l is the length of the rod. From these equations the constants L_1, L_2, etc., can be eliminated and a relation between $c_1 l$ and $c_2 l$ results, from which, since one relation already exists between $c_1 b$ and $c_2 b$, the quantities $c_1 l$ and $c_2 l$ can be found, and thus the normal functions can be determined.

So far this is much the same as the usual theory based on (41), the main difference being that c_1 and c_2 are not now equal, as they are in the simpler theory.

When, however, $p > \pi V_n/2b$, the p-hyperbola does not meet the curve β. At the critical value the point of intersection suddenly switches from the curve β to the curve ω_1, and, as p increases, the second point of intersection travels along this curve from $V=\infty$ till, at $p=\infty$, it reaches the asymptote where $V=V_n$. But the hyperbola also meets ω_2 when $p > 3\pi V_n/2b$ and goes on meeting more and more branches as p increases by intervals of $\pi V_n/2b$.

The simplest problem to solve is that of a rod pinned at both ends. The end conditions are

$$W=0, \quad M=0, \quad \text{at both ends.}$$

If $2a$ is the length of the rod and x is measured from the middle, both W and M are even functions for symmetrical modes and odd functions for asymmetrical modes. So it suffices to take

$$W=H \cos c_1 x, \quad M=K \cos c_1 x \quad \text{for even modes,}$$

and

$$W=H \sin c_1 x, \quad M=K \sin c_1 x \quad \text{for odd modes,}$$

where c_1 is obtained from the curve α only. All the boundary conditions can be satisfied by

$$\cos c_1 a=0 \quad \text{(even modes),}$$

or

$$\sin c_1 a=0 \quad \text{(odd modes).}$$

Thus the whole of the normal modes are obtained, for this case, from the single curve α. All wave-lengths from ∞ to 0 are included and, since infinitely short waves are needed to represent discontinuities, the problem of a pinned-pinned rod, whatever the initial conditions, can be completely solved by means of the curve α alone.

Let us next consider the same rod clamped at both ends, so that

$$W=0, \quad \frac{\partial W}{\partial x}=0, \text{ at } x=\pm a.$$

For this case, so long as $R_1^2 V^2/V_n^2 < \pi^2/4$, the value of W is

$$W=L_1 f_1 \cos c_1 x + L_3 f_3 \cosh c_2 x$$

and

$$\frac{\partial W}{\partial x}=-L_1 c_1 f_1 \sin c_1 x + L_3 c_2 f_3 \sinh c_2 x.$$

Thus the boundary conditions give

$$c_1 \tan c_1 a = -c_2 \tanh c_2 a.$$

This must be solved by means of the characteristic curves α and β. There are, of course, an infinite number of solutions. The greatest value of c_2b on β is $\pi/2$, and the corresponding wave-length $2\pi/c_1$ from α is $0{\cdot}327b$.

Observe that, by means of the first boundary condition, W can be expressed thus

$$W=H\left\{\frac{\cos c_1x}{\cos c_1a}-\frac{\cosh c_2x}{\cosh c_2a}\right\}.$$

If c_2a is a big number, either because c_2 is big or because a is big, then the second term in W has no appreciable value except near the ends of the rod. Thus, for c_2a and c_2x both big, we may write

$$W=H\left\{\frac{\cos c_1x}{\cos c_1a}-e^{c_2(a\pm x)}\right\},$$

the plus sign being correct where x is negative and the minus sign where x is positive.

This shows a feature which is true under all conditions, in the present theory as well as in the usual theory of vibrations, namely, that the hyperbolic functions of x that occur in the normal functions have, as a rule, no significant values except near the ends of the rod. That is, except near the ends of the rod, the normal functions are almost completely represented by the terms $\cos c_1x$ and $\sin c_1x$, the hyperbolic terms being dragged in merely to satisfy end conditions. The bigger c_1 is the more true is the last statement, and a big c_1 corresponds to short waves. That being so, it follows that only the first few modes—those of long wave-length—are seriously affected by the end conditions, except near the ends, and no hyperbolic terms occur in the expressions for the vibrations of a rod of infinite length.

When $R_1{}^2V^2/V_n{}^2>\pi^2/4$ and $<9\pi^2/4$ our equations give

$$W=L_1f_1\cos c_1x+L_3f_3\cos c_3x,$$

where c_3 is derived from the curve ω_1 and c_1 still from α. Then the equation for the c's is, for the clamped rod,

$$c_1\tan c_1a=c_3\tan c_3a.$$

This gives such an erratic series of values for c_1a that we are driven to the conclusion that the curves ω_1, ω_2, etc., do not play any part in the *primary* oscillations of the rod. To understand what is meant by primary here it must be recognized that the flat strip, of depth $2b$ and length $2a$, has a doubly infinite set of possible vibrations. There are wave-lengths in the z-direction as well as in the x-direction and there are fundamental vibrations with *overtones* for both directions. By primary then I mean those modes which contain no overtones in the z-direction but embrace all the overtones in the x-direction.

When $R_1{}^2V^2/V_n{}^2>9\pi^2/4$ a new trouble arises in that there are too many points of intersection for our few end conditions to cope with, The final conclusion then is that, when the p hyperbola no longer meets the curve β, the primary normal functions consist only of the terms $\cos c_1x$ and $\sin c_1x$ obtained from the curve α. This means that only one condition can be satisfied at each end of the rod now, and we have to decide what is the best form this end condition should take for the three different conditions of an end: (1) pinned; (2) clamped; (3) free. This may seem very inaccurate but it must be recalled that we failed to satisfy the exact conditions even for the slowest modes, and it is likely that the reduction in the number of constants for the quicker modes brings about no greater inaccuracy.

For consistency the single end condition for any type of end fixing must (1), agree with the corresponding pair of conditions at the critical value of $p(\pi V_n/2b)$, and (2), must be right at $c_1=\infty$. Now the equations (236) to (244) refer to the critical mode, and from these we see that

$$W=0, \quad M=0, \quad F=\text{finite}.$$

Next, to make comparisons when $c_1\to\infty$, we turn to (147 *a*) and (148 *a*). Here we see that M is of lower order than F, being of the order bF, and consequently, at a free end, the usual two conditions

$$F=0, \quad M=0,$$

are sufficiently covered by

$$F=0.$$

In the same way we can show that the two conditions for a clamped end

$$W=0, \quad \frac{\partial W}{\partial x}=0,$$

are sufficiently covered by

$$\frac{\partial W}{\partial x}=0.$$

For a pinned end the two conditions

$$W=0, \quad M=0,$$

are sufficiently covered by

$$M=0.$$

Conclusions.

The usual theory of the transverse vibrations of thin rods [equation (41)] is not seriously affected by the results in this paper so long as the wave-length is a fairly big multiple of the depth of the rod. Equation (41) is, in fact, reasonably good if the wave-length is greater than twenty times the depth of the rod (see equation (127)). But for small wave-lengths the errors become very big and the wave-form velocity for infinitely short waves is absurdly wrong in the usual theory, since only at $V=\infty$ does λ become zero, whereas the exact equations give $\lambda=0$ when $V_2=0{\cdot}86\,V_n$, *i. e.*, $V=0{\cdot}928\,V_n$. Now, when a purely local impulse is applied to a thin rod, the effect of this is transmitted along the rod by the progressive waves into which the normal modes can be analysed;

and a discontinuity in a Fourier or any similar series can be represented only by infinitely short waves. In so far as there is a wave front in the transfer of an impulse along a thin rod this is formed by the short waves, and the impulse or energy cannot travel faster than these short waves. My theory is that the greatest speed at which a transverse impulse can be transmitted along a thin rod is the above 0·928 $V_n(\sigma=0 \cdot 3)$, whereas equation (41) makes this impulse travel—in a diluted form, of course—with an infinite velocity. It can be shown that a *longitudinal* impulse can be transmitted with any velocity from 0·928 V_n to V_D.

In his celebrated ‘ Analytical Theory of Heat ’ (Chap. IX. Section IV. in Freeman's translation), Fourier, assuming the accuracy of equation (41), found an expression for the state of an infinite free rod at any time t after it was set free. As a piece of mathematics it is very wonderful, and is, of course, a correct deduction from (41). But as physics it fails for the reason that has just been stated, namely, that the impulses travel with infinite velocity in his results : a displacement at $x=0$ at time $t=0$ gives rise to a displacement at $x=a$ as soon as $t>0$, where a is any length whatever.

The same criticism applies to any result worked out by means of the normal functions of (41). Thus, according to this equation, a blow at one end of a finite rod produces an *immediate* effect at the other end, a result which is obviously absurd to a physicist.

Seismologists seem to be in accord that their instruments pick up, in turn, from a distant earthquake : (1) a dilatation wave ; (2) a distortion wave ; (3) skin waves. The first two do not get through as distinct waves in a thin rod.

We ought to try to answer the question raised a few pages back ; how can the oscillations of the type represented by the characteristic curves ω_1, ω_2, etc., be generated in the rod ? The answer is that there are two ways, namely (1) by impulses applied to the sides of the rod (the flat faces of the strip), and (2) by body forces, which we may have no physical means of applying but which the equations of elasticity are nevertheless capable of taking into account.

LXXX. *Some New Values for the Exponential Integral.*

By C. A. Coulson and W. E. Duncanson [†].

[Received July 20, 1942.]

1. *Introduction.*

The exponential integral is defined by the formula

$$\mathrm{Ei}(x)=\int_{-\infty}^{x} \frac{e^t}{t}\,dt. \qquad (1)$$

This function is repeatedly needed in the evaluation of integrals concerned with nuclear structure, molecular structure and flow of heat. Extensive tables are now available for a large range of x. We may refer to these as the B.A.$^{(1)}$ and W.P.A.$^{(2)}$ tables. They are so complete and detailed over the range $0<x<15$ that, in the words of the editor of the W.P.A. tables, "these present tables go a long way toward placing these functions in the class of known functions." The same is true for negative values of x which occur almost as frequently as positive values, and we shall find it convenient to write

---

[†] Communicated by the Authors.

$$\mathrm{Ei}^*(x) = -\mathrm{Ei}(-x) = \int_x^\infty \frac{e^{-t}}{t}\,dt. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

There are, however, no tables whatever of $\mathrm{Ei}(x)$ for values of x above 20 (the W.P.A. tables cover the range $0<x<15$), and none of $\mathrm{Ei}^*(x)$ in the range $15<x<20$. In fact, the only tables of $\mathrm{Ei}^*(x)$ for $x>20$ are those of Akahira$^{(3)}$ for $20<x<50$, which are given only to five figures. In the course of some quantum-mechanical studies we have required values of this integral to greater accuracy than this, and we have therefore thought it worth while to publish some of our numerical results.

There is, indeed, a particular reason for describing our work. Previous calculations for $x<15$ have been based on the series expansions

$$\mathrm{Ei}(x) = \gamma + \log x + \sum_{n=1}^{\infty} \frac{x^n}{n(n!)} \quad (\gamma = \text{Euler's constant}). \quad . \quad . \quad (3)$$

$$\mathrm{Ei}^*(x) = -\gamma - \log x + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}x^n}{n(n!)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Calculations for $x>15$ have depended on the divergent formulæ

$$\mathrm{Ei}(x) = \frac{e^x}{x} \sum_{n=0}^{\infty} \frac{n!}{x}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$\mathrm{Ei}^*(x) = \frac{e^{-x}}{x} \sum_{n=0}^{\infty} \frac{(-1)^n n!}{x^n}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

None of these formulæ is particularly convenient in the range $15<x<25$, especially in the case of $\mathrm{Ei}^*(x)$. Thus Glaisher$^{(4)}$, in his early tables, found it necessary to use powers of x up to x^{75} in (4) in order to get $\mathrm{Ei}^*(20)$ to two significant figures. Similarly when using the asymptotic formulæ it was only possible to get $\mathrm{Ei}^*(20)$ to six figures; after this point the series started to diverge. It is true that Airey$^{(5)}$ has quite recently shown that these asymptotic expansions can be used to give more figures than Glaisher obtained. But they are still cumbersome and we believe the following method to be much more efficient in calculating this integral to high accuracy.

2. *Formulæ for Calculation and Interpolation.*

It is not difficult to show (see below) that if $|h|<x$, then

$$\mathrm{Ei}(x+h) - \mathrm{Ei}(x) = \frac{e^{x+h}}{x} \sum_0^\infty \left(\frac{-h}{x}\right)^n \kappa_n(h) \quad . \quad . \quad . \quad . \quad . \quad (7)$$

$$= \frac{e^x}{x+h} \sum_0^\infty \left(\frac{h}{x+h}\right)^n \mu_n(h) \quad . \quad . \quad . \quad . \quad (8)$$

$$\mathrm{Ei}(x)-\mathrm{Ei}(x-h)=\frac{e^{x-h}}{x}\sum_{0}^{\infty}\left(\frac{h}{x}\right)^{n}\mu_{n}(h) \tag{9}$$

$$=\frac{e^{x}}{x-h}\sum_{0}^{\infty}\left(\frac{-h}{x-h}\right)^{n}\kappa_{n}(h) \tag{10}$$

$$\mathrm{Ei}^{*}(x)-\mathrm{Ei}^{*}(x+h)=\frac{e^{-(x+h)}}{x}\sum_{0}^{\infty}\left(\frac{-h}{x}\right)^{n}\mu_{n}(h). \tag{11}$$

$$=\frac{e^{-x}}{x+h}\sum_{0}^{\infty}\left(\frac{h}{x+h}\right)^{n}\kappa_{n}(h) \tag{12}$$

$$\mathrm{Ei}^{*}(x-h)-\mathrm{Ei}^{*}(x)=\frac{e^{-x+h}}{x}\sum_{0}^{\infty}\left(\frac{h}{x}\right)^{n}\kappa_{n}(h) \tag{13}$$

$$=\frac{e^{-x}}{x-h}\sum_{0}^{\infty}\left(\frac{-h}{x-h}\right)^{n}\mu_{n}(h), \tag{14}$$

where $\mu_n(h)$ and $\kappa_n(h)$ are functions of h satisfying the recurrence relations

$$n\mu_{n-1}=h(\mu_n+1), \quad \mu_0=e^h-1, \tag{15}$$

$$n\kappa_{n-1}=h(1-\kappa_n), \quad \kappa_0=1-e^{-h}. \tag{16}$$

Expressed in convergent power series, these are

$$\mu_n(h)=h/(n+1)+h^2/(n+1)(n+2)+h^3/(n+1)(n+2)(n+3)+\ldots, \tag{17}$$

$$\kappa_n(h)=h/(n+1)-h^2/(n+1)(n+2)+h^3/(n+1)(n+2)(n+3)-\ldots. \tag{18}$$

The functions μ and κ are related to the incomplete gamma function. In fact, since

$$\Gamma(n, h)=\int_0^h t^{n-1}e^{-t}\,dt,$$

therefore

$$\mu_n(h)=h^{-n}e^h\Gamma(n+1, h) \tag{19}$$

$$\kappa_n(h)=-\mu_n(-h)=(-1)^{n+1}h^{-n}e^{-h}\Gamma(n+1, -h). \tag{20}$$

These various formulæ enable us to write down μ_n and κ_n very simply if tables of the incomplete gamma function are available.

We do not need to prove all these relations here. Consider, for example, the relation (11). From the definition of $\mathrm{Ei}^*(x)$ given in (2),

$$\mathrm{Ei}^{*}(x)-\mathrm{Ei}^{*}(x+h)=\int_x^{x+h}\frac{e^{-t}}{t}\,dt=e^{-x}\int_0^h\frac{e^{-u}}{x+u}\,du$$

$$=\frac{e^{-x}}{x}\sum_{n=0}^{\infty}(-x)^{-n}\int_0^h e^{-u}u^n\,du, \text{ if } |h|<x,$$

$$=\frac{e^{-(x+h)}}{x}\sum_{n=0}^{\infty}\left(\frac{-h}{x}\right)^{n}\mu_n(h),$$

where
$$\mu_n(h)=h^{-n}e^h\int_0^h u^n e^{-u}\,du.$$

This gives (11) and (19). Integration by parts soon yields (15) and (17). In a similar way the other relations and the properties of $\kappa_n(h)$ are all obtained.

3. *The Present Tables.*

In this paper we give, in Table I., the values of Ei(x) and Ei*(x) between $x=15$ and $x=50$ at unit intervals. Considerations of space make it impossible to give a tabulation on the scale of the W.P.A. tables. Nor have we concerned ourselves with the range $x>50$ since the asymptotic

TABLE I.

Values of Ei(x) and Ei*(x).

x	Ei(x)		Ei*(x)	
15	2·349 558 525	10^{5}	1·918 627 892	10^{-8}
16	5·955 609 987	10^{5}	6·640 487 250	10^{-9}
17	1·516 637 894	10^{6}	2·306 431 990	10^{-9}
18	3·877 904 331	10^{6}	8·036 090 345	10^{-10}
19	9·950 907 251	10^{6}	2·807 829 097	10^{-10}
20	2·561 565 267	10^{7}	9·835 525 290	10^{-11}
21	6·612 718 636	10^{7}	3·453 201 267	10^{-11}
22	1·711 446 713	10^{8}	1·214 937 896	10^{-11}
23	4·439 663 698	10^{8}	4·282 684 795	10^{-12}
24	1·154 115 392	10^{9}	1·512 305 894	10^{-12}
25	3·005 950 907	10^{9}	5·348 899 755	10^{-13}
26	7·842 940 992	10^{9}	1·894 685 886	10^{-13}
27	2·049 649 712	10^{10}	6·720 637 435	10^{-14}
28	5·364 511 859	10^{10}	2·386 941 512	10^{-14}
29	1·405 991 958	10^{11}	8·487 759 779	10^{-15}
30	3·689 732 094	10^{11}	3·021 552 011	10^{-15}
31	9·694 555 759	10^{11}	1·076 767 039	10^{-15}
32	2·550 043 566	10^{12}	3·840 961 801	10^{-16}
33	6·714 640 184	10^{12}	1·371 384 348	10^{-16}
34	1·769 803 724	10^{13}	4·900 676 118	10^{-17}
35	4·669 055 014	10^{13}	1·752 705 939	10^{-17}
36	1·232 852 080	10^{14}	6·273 339 010	10^{-18}
37	3·257 988 999	10^{14}	2·247 020 698	10^{-18}
38	8·616 388 200	10^{14}	8·054 106 914	10^{-19}
39	2·280 446 200	10^{15}	2·888 779 301	10^{-19}
40	6·039 718 264	10^{15}	1·036 773 261	10^{-19}
41	1·600 664 914	10^{16}	3·723 166 776	10^{-20}
42	4·244 796 092	10^{16}	1·337 790 881	10^{-20}
43	1·126 348 290	10^{17}	4·809 496 557	10^{-21}
44	2·990 444 711	10^{17}	1·729 959 874	10^{-21}
45	7·943 916 028	10^{17}	6·225 690 809	10^{-22}
46	2·111 342 388	10^{18}	2·241 531 760	10^{-22}
47	5·614 329 680	10^{18}	8·074 197 843	10^{-23}
48	1·493 630 213	10^{19}	2·909 664 190	10^{-23}
49	3·975 442 748	10^{19}	1·048 981 164	10^{-23}
50	1·058 563 690	10^{20}	3·783 264 030	10^{-24}

formulæ (5) and (6) cover this range satisfactorily, and such large values of the argument are not often required. Our calculations were carried through using eleven significant figures, and the final values are given to ten. The last figure may therefore occasionally be in error by not more than ±2 units.

Interpolation between our values of Table I. is easily made if tables of the incomplete gamma function are available. But Pearson's tables (6) are scarcely accurate enough for values of $h>0\cdot1$, and we have therefore shown, in Table II., values of the auxiliary functions $\mu_n(h)$ and $\kappa_n(h)$ for $h=0\cdot1$, $0\cdot2$, $0\cdot3$, $0\cdot4$ and $0\cdot5$. These are sufficiently accurate to give ten significant figures in the final result. There is no need to give tables for $h=0\cdot6$, etc., since the basic formulæ (7)–(14) allow us to work either upwards or downwards from the nearest integer, and convergence is most rapid when h is kept small. With these tables interpolation to the first place of decimals in x is performed in a few minutes.

Thus to calculate Ei(20·2) we should use (7) in the form

$$\mathrm{Ei}(20\cdot2)-\mathrm{Ei}(20)=\frac{e^{20\cdot2}}{20}\sum_0^\infty\left(\frac{-0\cdot2}{20}\right)^n\kappa_n(0\cdot2).$$

Tables of the exponential function are available (7), so that all the required quantities are known. On account of the rapid decrease in $(0\cdot2/20)^n$ as n increases, only terms up to κ_4 are required to give an accuracy of 1 in 10^{10}.

Interpolation with $h<0\cdot1$ is easily done from Pearson's table, and only two or three terms are required. If Pearson's tables are not available it is a simple matter to determine the auxiliary functions to sufficient accuracy, and indeed this may sometimes be the easiest way (see Section 5 for a worked example).

4. *Details of Calculation of Tables.*

It is desirable to explain here the calculation and checking of Table I. First the table of Ei(x) was built stage by stage. The value of Ei(15) is given in the W.P.A. tables, and from this successive values at $x=16$, 17 . . . 50 were calculated, using (8). Each one was then checked independently, using (7). Further, at $x=30$, 40 and 50 an extra check was made by calculating directly by the asymptotic formula (5).

The values of Ei*(x) were obtained in the opposite way. Ei*(50) was calculated from the asymptotic formula (6) and then successively smaller values of x, $x=49$, 48, . . . 15, were taken, using (13). Each value was then checked independently by using (14). A further check was obtained by direct calculation at $x=40$, 30 from (6) and at $x=15$ by comparing with the value given by the W.P.A. and B.A. tables. No errors were detected. The first five figures of Ei*(x) were also checked where possible against the table of Akahira (3), and no errors were detected in this latter. Exponentials were taken from the W.P.A. tables of e^x (7), and powers of x from Barlow's tables.

TABLE II.

Values of the Auxiliary Functions $\kappa_n(h)$ and $\mu_n(h)$.

h =	0·1	0·2	0·3	0·4	0·5
κ_0 =	0·095 162 581 96	0·181 269 246 92	0·259 181 779 32	0·329 679 953 964	0·393 469 340 287
κ_1 =	0·048 374 180 4	0·093 653 765 4	0·136 060 735 6	0·175 800 115 09	0·213 061 319 43
κ_2 =	0·032 516 393	0·063 462 346	0·092 928 429	0·120 999 424 6	0·147 754 722 3
κ_3 =	0·024 508 2	0·048 064 81	0·070 715 71	0·092 504 316	0·113 471 666
κ_4 =	0·019 671	0·038 703 8	0·057 123 9	0·074 956 85	0·092 226 67
κ_5 =	0·016 4	0·032 404	0·047 935	0·063 039 4	0·077 733 3
κ_6 =	0·014	0·027 9	0·041 30	0·054 408	0·067 200
κ_7 =		0·024	0·036 3	0·047 9	0·059 19
κ_8 =			0·03	0·043	0·052 9
κ_9 =				0·03	0·048
κ_{10} =					0·04
μ_0 =	0·105 170 918 08	0·221 402 758 16	0·349 858 807 58	0·491 824 697 641	0·648 721 270 700
μ_1 =	0·051 709 180 8	0·107 013 790 8	0·166 196 025 3	0·229 561 744 10	0·297 442 541 40
μ_2 =	0·034 183 615	0·070 137 908	0·107 973 502	0·147 808 720 5	0·189 770 165 6
μ_3 =	0·025 508 46	0·052 068 62	0·079 735 02	0·108 565 404	0·138 620 994
μ_4 =	0·020 338	0·041 372 4	0·063 133 6	0·085 654 04	0·108 967 95
μ_5 =	0·016 9	0·034 310	0·052 226	0·070 675 5	0·089 679 5
μ_6 =	0·015	0·029 3	0·044 52	0·060 132	0·076 154
μ_7 =		0·026	0·038 8	0·052 3	0·066 15
μ_8 =			0·03	0·046	0·058 5
μ_9 =				0·04	0·052
μ_{10} =					0·05

At first sight it may be questioned whether this method of successive addition would result in some considerable error accumulating. This is not the case if we work towards larger values of x in calculating the table of $\mathrm{Ei}(x)$ and to smaller values in calculating $\mathrm{Ei}^*(x)$. These integrals change by a factor of 10 when x increases by an amount between 2 and 3, so that when x increases or decreases by unity in the two cases the greater part of each successive new value of Ei or Ei* arises from the terms on the right-hand side of equations (7)–(14); these terms may easily be calculated as accurately as we desire. This fact, and the independent back-checking described above, are sufficient guarantee that no cumulative error has appeared.

5. *A Worked Example.*

A worked example will illustrate our method of interpolation. Let us suppose that it is required to find $\mathrm{Ei}^*(20{\cdot}0234)$. For this purpose we use equation (11) with the values $x=20$, $h=0{\cdot}0234$, and therefore $h/x=0{\cdot}00117$.

Direct substitution of h in (17) gives, for $h=0{\cdot}0234$

$$\mu_4(h)=0{\cdot}004{,}698{,}3.$$

The recurrence relations (15) next give

$$\mu_3=0{\cdot}005{,}877{,}5, \quad \mu_2=0{\cdot}007{,}845{,}845, \quad \mu_1=0{\cdot}011{,}791{,}796{,}3$$

and

$$\mu_0=0{\cdot}023{,}675{,}928{,}035.$$

It is easier to use the recurrence relations in this order rather than in the reverse. Also it allows us to verify the accuracy of the various calculated values by comparing μ_0 with the correct value as given by (15), using the tables (7).

Substituting in (11) we discover that we only require μ_0, μ_1 and μ_2 to get the maximum possible accuracy, and thus, from Table I,

$$\mathrm{Ei}^*(20{\cdot}0234)=\mathrm{Ei}^*(20)-2{\cdot}382{,}165{,}55\ 10^{-12}$$

$$=9{\cdot}767{,}213{,}4\ 10^{-12}, \text{ correct to eight figures.}$$

An independent check is found from (12). In this instance the two results only differed by 2 in the next decimal place beyond that quoted above.

Had we desired to calculate $\mathrm{Ei}^*(20{\cdot}1234)$ we should first calculate $\mathrm{Ei}^*(20{\cdot}1)$ by the use of Table II., as described in Section 3, and then use equation (11) with $x=20{\cdot}1$ and $h=0{\cdot}0234$. The two stages would be quicker than one single direct determination taking $x=20$, $h=0{\cdot}1234$.

6. *Conclusion.*

The method used above can be extended quite easily to deal with the sine and cosine integrals. For these latter are merely combinations of exponential integrals with purely imaginary argument. The development of formulæ analogous to (7)–(14) is quite straightforward and we shall not reproduce the results here, particularly as excellent tables of $\mathrm{ci}(x)$ and $\mathrm{si}(x)$ are already available (2).

References.

(1) British Association, 'Mathematical Tables,' i. London, 1931.
(2) 'Tables of Sine, Cosine and Exponential Integrals,' Works Project Administration, National Bureau of Standards, 1940.
(3) Akahira, Sci. Pap. Inst. Phys. Chem. Res., Tokyo, June 1929.
(4) Glaisher, Phil. Trans. Roy. Soc. clx. p. 367, 1870.
(5) Airey, Phil. Mag. xxiv. p. 521, 1937.
(6) Pearson, 'Tables of the Incomplete Gamma Function,' H.M. Stationery Office, 1922.
(7) 'Tables of the Exponential Function,' Works Project Administration, National Bureau of Standards, 1939.

University College,
Dundee.

University College, London.
c/o University College of North Wales,
Bangor.

LXXXI. *Dimensions and the Facts of Measurement*[*].

By Norman Campbell[†].

[Received August 6, 1942.]

Introduction and Summary.

If this paper is to have any value its purpose must be clearly understood.

Disputants who fail to agree are often advised to define their terms. Formal definitions are inappropriate to science; the meaning of a scientific term is best conveyed by a statement of the facts that it is used to describe. Accordingly the familiar advice, translated into words appropriate to discussions about dimensions becomes :—State the facts from which such terms as " measure," " magnitude," " unit," " dimensions " derive their meaning. Formally, my object is merely to perform that task in the hope that disputants may thereby be helped to settle their differences.

If the task were as simple as it sounds, there would be no need to perform it; no competent physicist is really ignorant of the facts. The need for the task arises because the original meaning of a term often arises from very simple facts; it is used to describe more complex facts; it thereby changes its meaning more or less, and may acquire two meanings, each appropriate only to a different collection of facts. Further, conventions may arise whereby for certain purposes some of the facts are ignored or even distorted; in respect of the meaning of terms such conventions have the full force of true facts.

---

[*] Dr. Campbell has made this contribution at the request of one of the Editors (A. F.). In spite of Dr. Campbell's contention that such an article was hardly necessary, since his views had been fully put forward in his two books, ' Physics ' (C.U.P. 1920) and ' Measurement and Calculation '(Longmans Green, 1928), it seemed to the Editors that a statement of the gist of the arguments advanced in these books would be very helpful to the readers who are following the discussion. Hence the request for this article.

[†] Communicated by the Author.

I believe that complications of this kind have arisen in respect of "dimensions." Accordingly I shall first state (§§ 1–5) the simple facts which, if they stood alone, would give "dimensions" a definite meaning and a possible utility. I then introduce those facts, each associated with a convention, that make the meaning of the term much less definite and its utility doubtful (1). I shall proceed from the facts associated with (in my opinion) the more desirable conventions to those associated with the less desirable, so that the exposition will assume an increasing air of controversy.

It is, of course, open to anyone to say that I have stated or (more probably) selected the facts wrongly, and that I have misunderstood the conventions; that, if the facts were selected or stated otherwise, all the conventions would be justified and the significance of "dimensions" fully vindicated; and that all controversies about dimensions would then be resolved finally in some definite sense. But I suggest that such objections will be helpful only if the objector actually states the facts and conventions on which he relies in somewhat the same form and, above all, does not ignore any of what I regard as the "complicating" facts.

It is difficult, perhaps undesirable, to avoid the raising of side issues, some of them controversial. In order that the main theme may be expounded without interruption, I have placed all the notes and comments at the end. They are as bulky as the text—and no less important.

1. Dimensions are undoubtedly connected with algebraic (2) laws. Let us then start by considering what such a law means, bearing in mind the law of density and the law of falling bodies,

$$m=\rho v, \qquad \qquad (1)$$

and

$$h=\tfrac{1}{2}gt^2. \qquad \qquad (2)$$

The facts represented by these laws are these. We examine a collection of systems, all alike in some feature K (the constitution of a liquid or the place where the observations are made), but differing in respect of two variable magnitudes x, y. We measure associated pairs of the variables x, y, and write down the resulting numerals in two columns. We then find that the numeral in the y column, or some simple function of this numeral alone, is proportional to the corresponding numeral in the x column. The observations as a whole are therefore characterized by some constant of proportionality C.

We repeat these observations on other collections (other liquids or other places); we find a similar proportionality, but a different value of C. C is then a variable magnitude (density or gravitational acceleration) measuring the feature K. By measuring two magnitudes x, y we have produced a third magnitude C, which may be a variable x or y in some other algebraic law (3).

The result we have attained may be expressed by a definition (4), *e. g.*, that density is the ratio of mass to volume. But like all scientific definitions, such a statement implies a proposition, namely, that mass and volume are proportional.

2. By the continued application of this process the number of magnitudes grows. But each step is *dependent* on the prior measurement of other magnitudes; in order that the process shall start, there must be at least one magnitude (not two magnitudes—see § 11 below) whose measurement is *independent* of the prior measurement of any other magnitude.

3. There is a form of measurement, independent in this sense. It arises from the discovery of a pair of mutually converse transitive asymmetrical relations (greater or less) generating a series, and of a process (addition) whereby any pair of members of the series can be combined to form a third that is greater than either of them. Having discovered the relations and the process, we build up a "standard series" thus. Assigning to a selected member of the series the numeral 1, we find another neither greater nor less than (5) this "unit," combine it with the unit, find a member neither greater nor less than the combination, and assign to it the numeral 2—and so on. Having proceeded so far, we have to inquire (*e. g.*) whether the member denoted by 2 combined with the member denoted by 3 is consistently greater or less than the member denoted by 1 combined with the member denoted by 4. If it is neither greater nor less consistently, we have established an independent process of measurement; any other member is measured by finding a member of the standard series that is neither greater nor less than it, and assigning to it the numeral denoting that member of the standard series.

Everyone recognizes that this process is involved in producing a set of weights, a calibrated scale, or a resistance box. The relation "neither greater nor less than" is different in each case—balancing, coincidence of ends, equivalence in an electric circuit; so is the process of addition—effectively rigid connexion, juxtaposition of ends, connexion in series. But the features enumerated above are common to all (6).

4. Independent measurement is applicable only to the comparatively small number of magnitudes that are additive in the relevant sense. The list includes the following:—Number (of a collection) (7), length, volume, angle, mass (*i. e.*, that which is measured by a balance and "weights"), force, period of a system (which need not be "periodic"; an hour-glass has a period in this sense), electrical resistance, conductance, capacitance, permittance, charge, current, potential difference. The measurement of all non-additive magnitudes, which include all "qualities," must be dependent.

5. These facts suggest that we may arrange magnitudes in a genealogical tree. At the base would be " fundamental " magnitudes (8) measured independently. Next would come " primary derived magnitudes " measured by a dependent process involving fundamental magnitudes as variables ; then " secondary derived magnitudes " involving at least one primary derived magnitude as a variable, and so on. Then, if we adopted one of the conventional definitions of dimensions, we might perhaps assign to each magnitude unique dimensions. Moreover, if science had developed logically (9), such a tree would represent its actual development. (I deliberately refrain from attempting to define " dimensions," because the definition should depend on the use that is to be made of the conception ; and that is a controversial matter.)

6. However, there are two major complications that make such a tree impossible. The first arises from the fact that " the same magnitude can be measured in different ways." Since in the last resort a magnitude can be identified only by the way in which it is measured, this statement requires further examination.

Let us take the example of electric current. We set up a circuit including a battery and, in series with it, a silver voltameter and a straight conductor with a magnetic needle near it. Measuring m, the mass of silver deposited in time-interval t, we find $m=C_1t$; measuring the torque M on the needle and its distance r from the straight conductor, we find $M=C_2/r$. We now vary the E.M.F. of the battery. We find again algebraic laws of the same form, but with different constants C_1, C_2 ; and we find that, for all values of the E.M.F., $C_1=\alpha C_2$, where α is constant. Moreover, we find that there is a magnitude I, characteristic of the circuits, that can be measured independently by a standard series based on equality and addition ; and that $C_1=\beta I$, where β is another constant.

We might regard α, β as secondary derived magnitudes ; but they differ from most magnitudes in being invariable ; they are the same for all systems that are characterized at all by C_1, C_2, I. (10). Given one of these three, we can always find the other two without further experimenting. In these circumstances we say that C_1, C_2, I are all the same magnitude (11). Then the place of this magnitude in the genealogical tree is no longer unique ; and if we base dimensions on the tree, a magnitude has several alternative dimensions.

7. But a new consideration enters. So long as there is only one way of measuring a magnitude, we must use that way, whatever its accuracy. When there are several ways, there is good reason for selecting the most accurate way, and ignoring all the others. If we do that, we can again associate each magnitude with a single method of measurement ; we can draw up again a unique tree (" mensuration tree," we shall call it), and assign to each magnitude dimensions based on that tree. The drawing of such a mensuration tree is one of the tasks that we ought to perform in proposing a consistent set of " units " for universal

acceptance (12). But the mensuration tree is likely to be completely different from the historical tree, and so will be the dimensions based on it.

8. We now turn to the second main complication. A very important part of physics consists in establishing theories, relating algebraic laws (13). In this part of the science all laws are supposed to be already known; how they came to be known, and how magnitudes involved in them are measured, is irrelevant. The distinction between the variables and the constants of an algebraic law vanishes; (1) and (2) become simply relations between three variables, all on the same footing. The result of theorizing is to produce a tree of propositions, starting from propositions in some way fundamental and proceeding to primary, secondary, . . . propositions in some way consequent on them. Some magnitudes appear in the fundamental propositions, others appear first in the derivative propositions of various degree. The first appearance is of peculiar significance; for it is of the nature of a definition. Accordingly, regarding only first appearances, we can draw a "theoretical" tree, resembling in form the historical tree and the mensuration tree; and on this tree again we can base dimensions.

9. The theoretical tree will usually resemble somewhat the historical tree, but may differ very widely from the mensuration tree. Thus in the mensuration tree all the fundamental magnitudes must be measurable independently; there is no reason why this should be so in a theoretical tree. Again, since accuracy is, at least in principle, a matter of fact, it may be possible to say that one particular mensuration tree is right and all others wrong. But theories are partly matters of opinion; there may be many theoretical trees, each with its advantages and disadvantages, and appealing differently to different people (14).

A few simple examples may illustrate usefully what has just been said. In practical geometry length must be fundamental and distance derived. In abstract geometry the order is generally reversed. In electrical theory charge is almost always fundamental, and Coulomb's law occurs very early in the tree; capacitance and resistance are derived, and the laws that they are additive of little significance (15). In practical electrical measurements, on the other hand, the additivity of resistance (and perhaps capacitance) is essential; one or both of these magnitudes will be fundamental. Charge will certainly be derived and Coulomb's law will not be indicated at all; for it is not the most accurate way of measuring charge.

10. These two complications are inevitable; they arise from facts or from the essential nature of our study of them. But another complication is less inevitable; it arises from a refusal, without adequate reason alleged, to recognize as magnitudes things that appear to have all the characteristics of magnitudes.

Thus, there is a law relating the volume (v) of geometrically similar solids to their maximum linear dimension (L). Both v and L are measurable independently, and the law

$$v = \alpha L^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

is precisely similar in form to (1). Since α differs for (*e. g.*) a sphere and a cube, α is a magnitude entirely analogous to ρ; it might be called the *shape* of the solid. But α is generally ignored; most text-books say that the dimensions of v are L^3. No reason is given—and I can discover no reason—why, if α is to be ignored, ρ should not also be ignored, and why then we should not say that volume has the dimensions of mass (16). This neglect of magnitudes characteristic of geometrically similar systems is the source of much confusion in attempts to expound the (perfectly correct) principle of dynamical similarity in terms of dimensions.

11. Part of the reason for this neglect may arise from similarity laws in which the two variables are magnitudes of the same kind. Thus there is a law that, among rectangles of the same shape, the longer side is proportional to the shorter; if we put this is the form $y = \beta x$, β is a magnitude which might be called the shape of the rectangle. The refusal to recognize this magnitude may be due to a belief that x and y must be measured in terms of the same "unit," and that therefore the value of β is independent of any choice of "units" (17). But there is no "must" about it; so long as the longer and shorter sides are distinguishable, we might decide to measure one in inches and the other in centimetres; if we changed our choice, and measured one in miles and the other in yards, β would change. However, it is true that, if this possibility of changing "units" (18) is any part of the definition of a magnitude, we should have to admit that the shape of a square is not a magnitude; for then x and y can be no longer distinguished (19).

12. We are thus led at last to consider the relation between "units" and dimensions, a relation that is sometimes thought to be so intimate that it ought to form the starting-point, rather than the conclusion, of such a discussion as this (20).

The laws underlying independent measurement are such that there is no limit to our choice of the unit member to which the numeral 1 is assigned. If, having hitherto chosen a member A, we transfer our choice to another B, greater or less than A, the numeral assigned to each member of the series will have to be changed by a constant factor. If algebraic laws are always written in the form $f_1(x) = C \, . \, f_2(y)$, each C will change by a constant factor when the unit of x or y is changed; and the value of this factor corresponding to any given change in the unit of x or y is wholly determined by the form of f_1 and f_2, and can be described in terms of the dimensions of C.

13. But there is no reason why algebraic laws should always be written in this form. A constant of proportionality can always be

resolved into a product of two (or more) constants. We might always write

$$f_1(x)=\mathrm{F}.\mathrm{C}'.f_2(g), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and adopt the convention that, when the system subject to the law changes, C′ is to be changed, but not F. This is actually done in the law (2); there is no difference in the experiments on which laws (1) and (2) are based that compels us to introduce the $\frac{1}{2}$ in (2) or forbids us to introduce (say) π into the right-hand side of (1). Having resolved C into a magnitude C′ and a "fixed constant" F, we may now adjust the law, when the unit of of x or y changes, by changing F and not C′; then the "unit" of the derived magnitude will not change at all with the units of the fundamental magnitudes. This is not a mere ideal possibility; it is actually used both by engineers and by theoretical physicists. There is therefore no necessary connexion between changes in the "units" of derived and of fundamental magnitudes; any discussion of dimensions that assumes that there is a necessary connexion must be inadequate.

The same considerations enter when we are proposing a set of "units" for general acceptance. Having drawn a mensuration tree, we may first choose units for the fundamental magnitudes measured independently. In connexion with each primary derived magnitude, we have another arbitrary choice; we may either assign a numeral to some system having that magnitude and thereby determine the fixed constant; or we may choose the fixed constant and thereby determine the numeral to be assigned to every system. Alternatively we may leave one of the fundamental units undetermined at the outset, and, when we come to a derived magnitude, choose both the fixed constant and the numeral to be assigned to some particular system having the derived magnitude. Then we have determined the numeral to be assigned to any system in respect of the fundamental magnitude and have in effect chosen its "unit." Thus, if mass and volume are fundamental magnitudes, we may leave the unit of volume undetermined and, when we come to law (1), choose that both the fixed constant and the density of water shall be 1; we have then fixed the "unit" of volume (21). Whichever alternative we choose, we have always as many arbitrary choices as there are magnitudes in our table.

14. In choosing between these alternatives we may reasonably be guided by a desire (*a*) to secure the greatest possible uniformity of the "units" in which experimenters express their results, (*b*) to facilitate arithmetic. Object (*b*) is best promoted by making each fixed constant a power of 10, including 1. Kelvin, in proposing his "absolute units," thought that the same choice promoted object (*a*), and that therefore as many "units" as possible should be determined by assigning values to fixed constants (22). He may have been right at the time; but now it is known that uniformity is best promoted by setting up standardizing

laboratories which interchange their material working standards. "Units" are then determined by the numerals assigned by agreement to these standards. But though, in order to attain uniformity, we assign numerals to material standards and not to fixed constants, the assignment need not be arbitrary; we may assign them so that the fixed constants are very approximately powers of 10. This is the solution proposed by the International Committee of 1935 (23); their only definite instruction, addressed to standardizing laboratories, was that, while the numerals assigned to existing standards should generally be changed as little as possible, they should be changed, if necessary, so as to make certain fixed constants that are now near powers of 10 even more nearly powers of 10.

15. When we are dealing with a theoretical tree, the objects corresponding to (*a*) and (*b*) are (*c*) that propositions in the tree are true when the "units" fixed by the standardizing laboratories are used, and (*d*) that there are no unnecessary symbols. In view of experimental error, (*c*) is strictly unattainable; and the practice of theoretical physicists shows that, even if (*c*) were attainable, (*d*) would be much more important.

One method of securing object (*d*) would be for writers of general treatises always to represent the fixed constants by algebraic symbols, at least on the first occasion when they state any equation. In the discussion of a particular problem, usually only some of the equations appear. If these are to be independent and non-redundant, the number of equations must be equal to the number of symbols that actually vary during the discussion of the problem. It is then permissible (but not necessary) to substitute 1 for any symbol that does not vary during the discussion, whether it be a fixed constant or a magnitude—so long as the public are clearly told what has been done in a manner that enables them to translate the propositions stated into terms of the standardized units. This condition is not always easy to fulfil. But it could be fulfilled if there were devoted to the devising of appropriate conventions the ingenuity that has hitherto been expended in inventing "theories of dimensions."

Notes.

(1) Perhaps I should answer my own questions. (Phil. Mag. xxxiii. p. 398, 1942.) In my view dimensions are positively useful only for the trivial object (1). For object (2) they are harmless; for the others definitely confusing.

(2) Hitherto I have used the term "numerical" in preference to "algebraic" laws, with the object of emphasizing the important part played by numerals in measurement and their distinction from numbers in the sense of § 4 below. Some readers may jib at the statement that proportionality is a relation between numerals and not between numbers. All that I am concerned to assert here is that, if proportionality is said to be a relation between "numbers," then "number" is used in a sense quite different from that in which it is used in § 4.

(3) An algebraic law may involve more than two variables; but in practice such laws are seldom established except by combining several laws, each involving only two variables. A law with two variables may generate more than one constant; thus the law $y = A_0 \sin \omega t$ generates the two constants A_0 and ω. These complications do not really affect the discussion.

(4) Other definitions of magnitudes imply other kinds of proposition. Thus the definition of kinematic viscosity as the ratio of viscosity to density obviously does not imply that, in every fluid, viscosity is proportional to density. I do not propose to discuss here what other kinds of propositions may be implied, or whether such definitions indicate that the measurement of one magnitude may be dependent on that of others in a manner other than due to their relation by an algebraic law. For, if there are other manners of dependence, they are equivalent *for our present purpose* to dependence arising from an algebraic law.

(5) Pairs of transitive asymmetrical relations are sometimes such that " neither greater nor less than " is a transitive symmetrical relation (equality). It is important to note that this is *not* true for the relations characteristic of any system of independent measurement other than that of number; for experimental error arises from the fact that " neither greater nor less than " is not transitive. But for brevity I shall sometimes call the pair of relations by the single name " equality." Again for brevity I say nothing about extending the standard series to magnitudes less than the unit; the process is tedious to describe, but its nature is obvious once it is understood how the series greater than the unit is established.

(6) Is there any other process by which a magnitude can be measured independently of the prior measurement of others? The question is often raised in connexion with apparent magnitudes such as probability or intensity of sensation. The answer depends on what restrictions are imposed on " measurement." I hold (but others differ) that, if " measurement " implies *all* the qualities that are essential to the importance of measurement in physics, then there is no other method of independent measurement.

(7) It may be argued plausibly that the only one of these magnitudes whose measurement is entirely independent of that of any other is a number, and that in measuring all the others we use numbers. This is a matter on which I have changed my opinion more than once; so I will express none here.

I have included in the list only those that are in some way relevant to the present discussion. There are, of course, several others that could certainly be included in the list, and some whose inclusion is doubtful. But it seems necessary to say something about one that is not included.

Time-interval, the magnitude t in (2), is related to period much as distance is to length. The time-interval between two events is the period of a system whose characteristic events are respectively simultaneous with the two events. The contrary is often stated; it is asserted that the measurement of time-interval is dependent on uniform velocity and/or length (see, *e. g.*, L. Silberstein, Phil. Mag. xxxviii. p. 382 (1919) and xxxix. p. 366 (1920). I simply cannot understand the basis of such assertions; it seems to me perfectly plain that time-interval can be, and usually is, measured as dependent on period and on nothing else, and that period is measured independently.

On the other hand, time, the uncontrollable variable, is in my opinion a theoretical conception that is not measurable at all.

(8) Hitherto I have used the terms " fundamental " and " derived," not for magnitudes, but for the methods of measurement here called independent and dependent. The present usage is probably better; but it must be observed that, in their present usage, fundamental and derived do not denote qualities

in isolation; they represent a relation; as we shall see, a magnitude that is fundamental in one aspect may be derived in another.

(9) This hypothesis is very far from the truth. Thus all photometric magnitudes are supposed to be based on the inverse-square law. But direct evidence for that law hardly exists.

(10) It is important to distinguish α and β from what are usually called "universal constants." This phrase is really a misnomer. If in all systems energy were proportional to characteristic frequency, h in the relation $E=h\nu$ would probably be ignored. It is so important just because the proportionality of energy to frequency in a special class of systems is so surprising; the significance of "universal constants" lies in their particularity.

(11) The question whether we should regard resistance and conductance as "the same magnitude" depends on the importance we attribute to dimensions. They are the same in the sense that, when we know one, we know the other; and they must appear in the same rank of the genealogical tree; but according to most, if not all, definitions of dimensions they have different dimensions.

(12) See, *e. g.*, N. Campbell, Proc. Phys. Soc. xlviii. p. 706 (1936). This paper illustrates well how widely a mensuration tree may differ from a historical tree.

(13) "Theory" is used here to denote an ordered set of propositions. The order usually results from the introduction of "theories" in a narrower sense, *i. e.*, propositions differing in character from those that they co-ordinate. I still believe that the distinction between laws and theories (in this narrower sense) expounded in my *Physics* is valid and important for many purposes. But it is not important here except in so far as it indicates that there may be variables that do not represent any measurable magnitude at all.

(14) I seem to remember a controversy in which the disputants, though professedly arguing about the dimensions of μ, turned out to be really arguing whether, in the instruction of engineering students, magnetic induction should be introduced first as the field in a special kind of slot or as that whose variation determines an E.M.F.

(15) There are many well-known treatises of Electricity and Magnetism that make no mention at all of the fact, all-important in practical measurement, that resistance is additive when resistors are placed in series and conductance when they are placed in parallel. Even more surprisingly these facts are often ignored in the discussions that arise from time to time concerning the meaning of Ohm's law. It is essential to the *present* significance of that law that E.M.F., current and resistance can all be measured independently.

(16) I have always found this example very effective in destroying the child-like faith which believes that any statement repeated through the generations from text-book to text-book is necessarily indisputable.

(17) Magnitudes resembling this β are sometimes called "pure numbers." The phrase seems to me very unfortunate; for these magnitudes are totally different from the number of a collection which plays such an important part in all measurement. It is *not* true that the number of a collection is independent of the unit; the number of an army is different according as the unit is a division or a single soldier.

(18) I put *units* here and elsewhere in inverted commas, because the term seems to imply the false doctrine that the system to which the numeral 1 is assigned must always differ importantly from other systems. That is so only when the "unit" of independent measurement is fixed by assigning the numeral 1 to some system identified as an individual, *e. g.*, the rotating earth or the standard kilogramme.

(19) These remarks are relevant to the interminable controversy about the dimensions of μ. According to Maxwell μ is the ratio of the fields in two

different kinds of slot. If so, it is said, μ must be no-dimensional. But others point out that, if the theory is expounded in another way, μ turns out to have definite dimensions. It is *not* true that, even on the Maxwellian view, μ is necessarily no-dimensional ; but there is some truth in the statement that, if Maxwellian permeability is dimensional, then the permeability of a vacuum does not exist.

(20) It seems permissible to point out a source of much of the existing confusion. International Committees are appointed at long intervals to fix "units." Believing that units and dimensions are closely associated, the members appointed feel bound to assume at short notice precise views on matters that they have never considered seriously. They naturally fall victims to the speciously simple doctrines of persons better versed in publicity than in accurate thought. During the ensuing discussions, these doctrines are found to be false—and also irrelevant to the decisions of the Committee. The errors are therefore never exposed authoritatively, and the subject is left in such a tangle that no text-book writer dares to touch it.

Is it too much to hope that, before the next International Committee meets, a subsidiary committee should be charged with defining precisely the questions that it has to decide and pointing out what generally accepted propositions are relevant to its decisions ? It would be an admission of dreadful incompetency if physicists had to admit that this duty could not be discharged.

(21) This process is sometimes regarded as the identification of a unit by specification rather than by demonstration, *i. e.*, by saying that it is such-and-such rather than that it is *that.* But this view is false. It is practically impossible to make a unit precisely to specification ; a law, valid for at least small variations, is always assumed. But the use of the law for fixing the unit does not imply that the measurement of the unit is to be always dependent. Thus we may fix the unit of resistance by the law of the Lorenz apparatus, but still measure other resistances in terms of that unit by combination in series and parallel.

(22) Since a mensuration tree must have at least one magnitude measured independently, not all "units" can be assigned by fixed constants. The inquiry how many have to be assigned otherwise is very intricate ; the view adopted by many (but apparently not all) of the members of the 1935 Committee, that exactly four "units" have to be assigned otherwise in an electrical system, is certainly not the only one that may be plausibly maintained. It is amusing to note that the "fourth unit" that appealed to those who held this view belongs to a magnitude ("permeability of space") which, on Maxwell's theory, is non-existent. (See § 11.)

Again not all standards can be maintained by interchange ; for some (*e. g.*, that of current) are not easily embodied in interchangeable apparatus. Accordingly the standardizing laboratories have to rely partly on agreed fixed constants.

Planck's proposal for "absolute units" was to fix both the fixed constant and the derived magnitude in laws involving "universal constants." But that proposal is entirely contrary to object (*a*).

(23) Since I have criticized some members of the Committee, perhaps I had better say that this decision appears to me eminently wise. But it was quite independent of the views on which it was apparently based.

LXXXII. *Electric Determination of Film-volume and Film-density.*

By H. Löwy, Dr.Phil.*

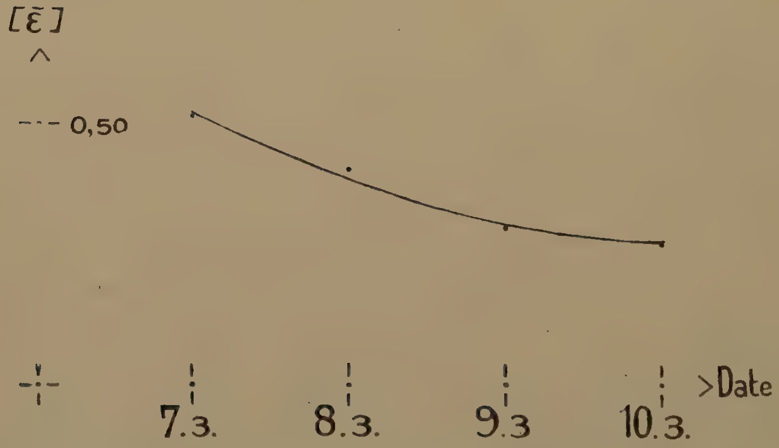
[Received June 12, 1939.]

In a previous paper ⁽¹⁾, I have deduced the fundamental electrodynamic hydrological equations under the supposition that the water is dispersed in the rock in form of drops. In reality, the water covers the rock-particles in the form of films. Therefore, the *matter-volume* v_w of the water-drops must be replaced by the *electrodynamic volume* ⁽²⁾ of the water-films, and the hydrologic equations become

Table I.

Date.	$\bar{\epsilon}$.	$[\bar{\epsilon}]$.
7.3	4·3	0·52
8.3	3·1	0·41
9.3	2·3	0·30
10.3	2·1	0·27

Fig. 1.



$$[\bar{\epsilon}]=V_w+v_c[\epsilon_c]. \quad \text{(1)}$$

$$v_c+v_a+V_w=1. \quad \text{(2)}$$

---

* Communicated by the Author.

$\bar{\epsilon}$ and ϵ_c designate the dielectric constants of the rock and the rock-substance. v_a is the volume of air in 1 c.cm. of the rock; v_c the volume (in 1 c.cm. rock) of the rock-particles which are *not* covered with water-films.

In the case when the water covers every rock-particle, these equations reduce to

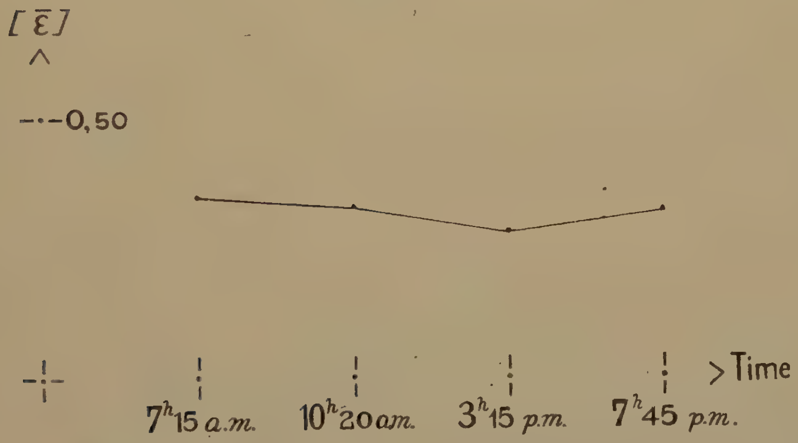
$$[\bar{\epsilon}]=V_w. \qquad (3)$$

$$v_a+V_w=1. \qquad (4)$$

TABLE II.

Time.	$\bar{\epsilon}$.	$[\bar{\epsilon}]$.
7h 15 A.M.	2·6	0·35
10h 20 A.M.	2·5	0·33
3h 15 P.M.	2·2	0·28
7h 45 P.M.	2·4	0·32

Fig. 2.



In this case, the film-volume (per c.cm.) v_f and the mean film-density d_f are given by the following simple formulæ

$$v_f=V_w-v_c. \qquad (5)$$

$$d_f=\frac{v_w d_w}{v_f}. \qquad (6)$$

v_c has another signification in (5) than in (1). It designates now the volume of *all* rock-particles contained in 1 c.cm. of the rock. We can now also write $v_c=1-p$, if p is the porosity of the rock.

v_w designates the matter-volume which the water, contained in 1 c.cm. of the rock, would fill in its normal state (water-density $d_w=1$).

In March 1939 I measured, at a certain point in the *Western Desert of Egypt*, the temporal variations of the dielectric constant ϵ of the soil after a rainy day (March 6). In Table I. and fig. 1 are given the values of ϵ and $[\epsilon]$ on the four following fine-weather-days (March 7 to 10). Table II. and fig. 2 contain the analogous variations during one fine-weather day (March 11). Such measurements may prove to be of interest for agriculturists in arid and semi-arid countries.

On March 7, I found $[\epsilon]=0{\cdot}52$ or, according to (3), $V_w=0{\cdot}52$; hence we obtain for v_c the value 0·50, that is a soil-porosity $1-v_c$ of 50 per cent., according to (5), the film-volume $v_f=0{\cdot}02$.

Designating by N the number of sand grains contained in 1 c.cm. of the soil, by r the radius of the grains, we have

$$N\,.\,\frac{4\pi}{3}r^3=v_c, \quad \ldots\ldots\ldots \tag{7}$$

and for $v_c=0{\cdot}50$ and $r=10^{-3}$ cm. (fine sand (3)) we find $N=10^8$.

Film-volume v_f and film-thickness b_f are connected by the equation

$$v_f=N\,.\,4\pi r^2\,.\,b_f \quad \ldots\ldots\ldots \tag{8}$$

In our case ($v_f=0{\cdot}02$) we obtain

$$b_f=1{\cdot}6\,.\,10^{-5}\text{ cm.},$$

a value which is in conformity with the film-thickness $b_f=1{\cdot}3\,.\,10^{-5}$ cm., measured by Earl Pettijohn (4) on Ottawa sand. The radius of the desert sand grains ($r=10^{-3}$ cm.) being much smaller than the radius of the Ottawa and other sands used by Pettijohn ($r=1{\cdot}7\,.\,10^{-2}$ to ($7{\cdot}9\,.\,10^{-2}$ cm.), our measurement confirms, over a wider range, Pettijohn's result that " the film-thickness is independent of the size of the solid particle."

From the value $[\epsilon]=0{\cdot}41$, measured on March 8, and from the values of $[\epsilon]$, measured on the following days, a negative value of the film-volume results, according to (5). During the 24 hours following the measurement of March 7, the desiccation of the soil has made such progress that the remaining films contain only " *solidified water* " (5). These water-films are no longer electrically conductive. In this stage of dryness the hydrologic equations must be replaced by the equations of H. A. Lorentz (6) for dielectric mixtures.

REFERENCES.

(1) H. Löwy, Phil. Mag. xxvii. p. 576 (1939).
(2) H. Löwy, Phil. Mag. xxvi. p. 453 (1938).
(3) C. A. Hogentogler, ' Engineering Properties of Soil ' (New York, 1937 : McGraw Hill Book Co.), chapter 1.
(4) Earl Pettijohn, Journ. Amer. Chem. Soc. xli. p. 477 (1919).
(5) C. A. Hogentogler, *l. c.* chapter 5, fig. 30.
(6) H. A. Lorentz, ' Theory of Elections,' § 126 (Leipzig, 1909 : Teubner).

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

LXXXIII. *Some Notes and Suggestions on the Teaching of Physics.*

By C. J. Smith, Ph.D., M.Sc., A.R.C.S., Lecturer in Physics, Royal Holloway College (University of London), Englefield Green, Surrey[*].

[Received July 30, 1942.]

In the following paper a plea is made for greater exactness in defining physical quantities and for greater attention to the correct use of algebraic signs. Examples are cited to illustrate some of the points raised. In addition, during the author's work as a teacher of physics, several well-known experiments have been improved and a short description of the improvements is given. Attention is also drawn to some changes which seem desirable in the matter of notation.

The Ideal Gas Equation.

If p is the pressure, v the volume per gram of an ideal gas, T the absolute temperature, then

$$pv = \mathscr{R}\mathrm{T}.$$

If V is the volume per mole, then

$$p\mathrm{V} = \mathrm{RT}.$$

Hence $\mathrm{R} = \mathrm{M}\mathscr{R}$, where M is the molecular weight of the substance. The advantage of using the equation $p\mathrm{V} = \mathrm{RT}$ is that R is a universal constant, viz., $8{\cdot}31 \times 10^7$ erg. mole.$^{-1}$ deg.$^{-1}$ K. This is the equation universally used by chemists: on the other hand physicists, especially in dealing with the principles of gas thermometry, find it convenient to consider one gram of a substance. It is therefore essential to make a clear distinction between the two gas constants $\mathscr{R}$ and R, the former of which is expressed in erg. gm.$^{-1}$ deg.$^{-1}$ K. (or any other consistent set of units). In printed work **R** is used for R, but this is inconvenient in written work and for teaching purposes.

Some authors have used r for $\mathscr{R}$, but this is not desirable, since r generally denotes a distance.

Some Remarks on the Use of Searle's Surface Tension Balance.

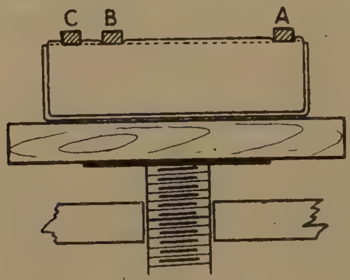
In using this well-known piece of apparatus, sufficient attention is seldom given to the cleanliness of the liquid (water) surface. A clean surface is rapidly obtained in the following way:—The glass crystallizing dish, which serves eventually as the container for the liquid and whose rim has been coated with clean paraffin wax, is cleaned in the usual way and then finally filled with water so that the water-surface stands above the top of the vessel (this top is shown dotted in fig. 1). Glass strips, A, B and C, each about 1·5 cm. wide and coated with wax, are then allowed to rest on the rim. The water surface is first swept with B; A is then removed to the position C, and the surface again swept with the "middle" bar. Three or four such sweeps are more than sufficient to clean the surface thoroughly.

---

[*] Communicated by the Author.

The balance having been equilibrated the vessel is moved slowly upwards—there should be no rotation of the vessel about a vertical axis—until the water surface just touches the suspended glass plate, the lower edge of which has been made horizontal. The rider is then moved until the plate breaks away from the surface when γ, the surface tension, may be calculated in the usual manner. No correction for buoyancy is necessary.

Fig. 1.



On Sentis' Method of Determining the Surface Tension of a Liquid.

The advantages of this method for determining the surface tension of a liquid, as well as its theory, have been given recently by Ferguson*. It is surprising to find, however, that in certain text-books the method is stated to require a knowledge of the "angle of contact" which the liquid makes with glass. This is the more remarkable since the method was first described in 1887†, and in the opening paragraph of the original paper we read "*Voici une méthode analogue qui est indépendante de cette hypothèse*" (que l'angle de raccordement est nul).

The author of the present paper has often wondered why, in text-books of practical physics and elsewhere, one is instructed "to draw down the capillary at its lower end": one can see at once that a smaller drop will be formed when such a tube is made, but now that a more complete theory has been given by Ferguson it seems a pity to constrict the tube in the above manner for the following reason:—When the capillary tube is constricted at its lower end, the formation of the drop is hindered if the viscosity of the liquid is appreciable, so that measurements may be made before truly steady conditions have been attained. The author has found that this difficulty is avoided if the tube is constructed as shown in fig. 2. A glass tube about 2 mm. in internal diameter is sealed on to the lower end of the capillary tube (0·5 mm. to 1 mm. in diameter), and then cut off so that its length is about 5 mm. Well-defined drops, easily measurable, are then formed very rapidly even when liquids with a high coefficient of viscosity are used, and reliable values for the surface tension are obtained.

---

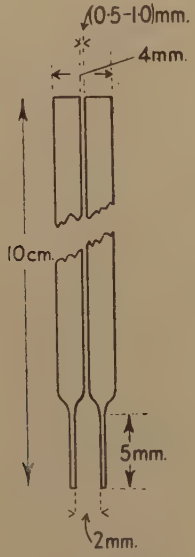
* Ferguson, Proc. Phys. Soc. liii. p. 554 (1941).

† Sentis, *J. de Physique*, vi. p. 571 (1887).

The Variation of the Saturation Vapour Pressure of a Liquid due to Curvature of its Surface.

The treatment of this problem in most text-books cannot be regarded as being anything like satisfactory on account of the fact that signs are confused and the assumptions made are not clearly stated. Moreover, it seems to the author that it is not desirable to deal with the problem when the surface is concave and then state that for convex surfaces, which are the important ones, the effect will be reversed in sign. The following method is therefore suggested:—Consider the arrangement shown in fig. 3 (*a*). The liquid is contained in a vertical tube so wide at the top that the liquid surface is plane, and so narrow below that a very small drop of liquid is formed at its lower end, as in Sentis' method of finding the surface tension of a liquid. The whole is surrounded by the vapour of the liquid only.

Fig. 2.



Let p_0 be the saturation vapour pressure of the liquid over the plane surface, and p_c that over the curved surface of the drop the radius of which is r. All heights, z, are measured from the tip of the drop, the total height of the liquid being Z.

Let p and $p+\delta p$ be the pressures in the vapour at heights z and $z+\delta z$, the corresponding vapour densities being σ and $\sigma+\delta\sigma$. Then, if g is the intensity of gravity,

$$dp=-\sigma g \,.\, dz,$$

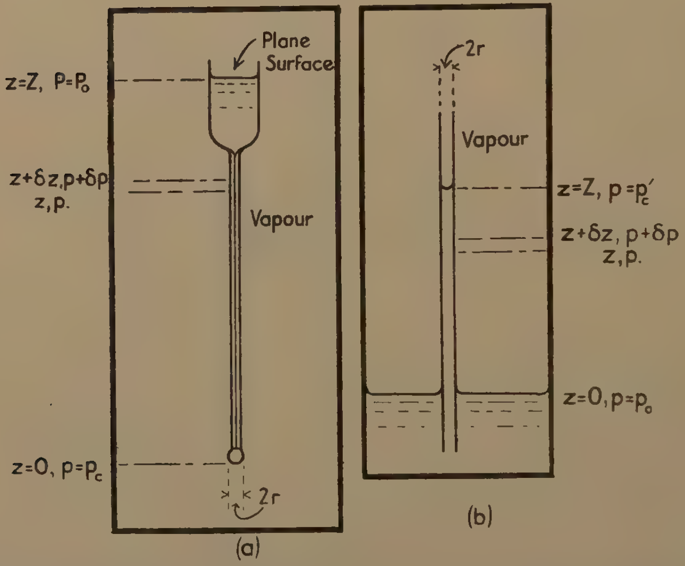
$$\therefore \qquad \frac{dp}{p}=-\frac{gM}{RT}dz,$$

where M is the molecular weight of the substance, R the gas constant for one mole, and T is the absolute temperature. In making this substitution for σ it is assumed that the vapour behaves as an ideal gas. On integrating this equation, T being assumed constant,

$$\log\frac{p_0}{p_c}=-\frac{gM}{RT}Z.$$

$$\therefore \qquad p_c>p_0.$$

Fig. 3.



If γ is the surface tension of the liquid, and we assume that its density ρ is constant,

$$p_0+g\rho Z=p_c+\frac{2\gamma}{r}.$$

$$\therefore \qquad \log\frac{p_0}{p_c}=\frac{M}{RT}\,\frac{1}{\rho}\left[p_0-p_c-\frac{2\gamma}{r}\right],$$

or

$$\log\frac{p_c}{p_0}=\frac{M}{RT\rho}\left[\frac{2\gamma}{r}-(p_0-p_c)\right].$$

Now p_0-p_c is small in comparison with $\frac{2\gamma}{r}$ and may be neglected.

$$\therefore \qquad \log\frac{p_c}{p_0}=\frac{M}{RT\rho}\cdot\frac{2\gamma}{r}.$$

Hence, at a convex liquid-vapour surface the saturation pressure of the vapour in equilibrium with that surface is greater than the saturation pressure of the vapour in equilibrium with a plane liquid-vapour surface.

To show that $p_c - p_0 > 0$ for a convex surface it is not necessary to make Z large. If it is small we may assume σ to be constant. Then

$$dp = -g\sigma\, dz,$$

$$p_c - p_0 = g\sigma Z,$$

which is positive. In actual practice $p_c - p_0$ is only appreciable when Z is large, *i. e.*, the radius of the drop is very small. The logarithmic formula must always be used under such conditions.

The argument for a concave surface follows at once from considerations of fig. 3 (*b*). With the same assumptions as before

$$\frac{dp}{p} = -\frac{gM}{RT}\, dz,$$

and

$$p_0 = p_c' - \frac{2\gamma}{r} + g\rho Z,$$

$$\therefore \quad \frac{RT}{M} \log \frac{p_0}{p_c'} = gZ = \frac{1}{\rho}\left[p_0 - p_c' + \frac{2\gamma}{r}\right],$$

where p_c' is the pressure of the saturated vapour over the concave surface.

The Coefficient of Viscosity of a Gas as Interpreted by The Elementary Kinetic Theory of Gases.

In this paragraph the writer desires to point out a very common fallacy which still occurs in some text-books and elsewhere. Consider a plane area A in a fluid in steady motion, the area being parallel to a plane xOy, in which there is no motion. At all other places the motion is parallel to this plane. If v is the velocity of the fluid at all points in A, the coefficient of viscosity, η, is defined by the equation

$$\frac{F}{A} = \eta \frac{dv}{dz},$$

where F is the force which the fluid above A exerts on the fluid in the plane A. This is Newton's hypothesis *.

To calculate a value for the coefficient of viscosity of a gas, let the direction of the flow of the gas be parallel to the plane xOy and consider a unit-plane AB, fig. 4 (*b*), parallel to this plane, *i. e.*, AB is a plane of unit area. The mass-velocity of the gas at any point in a plane z=constant is constant. Let v be the mass-velocity of the gas at any point in AB. Let us calculate the rate at which momentum due to the mass-velocity of the gas is being transferred across the unit-plane. It will be supposed that v increases upwards and assumed, with Joule, that molecules only cross the plane normally from above and below.

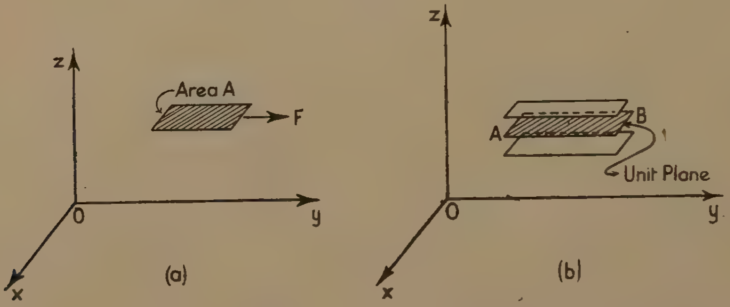
---

* *Cf.* Barr, 'Viscometry,' p. 3.

Those molecules which pass downward will have greater momentum due to their mass-velocity than those in the plane : those passing upward less. Since there is no mass-motion parallel to the z-direction, the number of molecules crossing the plane per second is the same in each direction. Hence there is a downward transfer of momentum in the y-direction due to molecules travelling parallel to the plane xOz.

The molecules passing across the plane will have come from different distances since colliding with another molecule. It will be supposed that those passing downwards have all come from a plane at a height λ above the unit-plane AB, where λ is the mean free-path. Their mass-velocity will be $v+\lambda\frac{\partial v}{\partial z}$. Now the velocity of any molecule in the plane AB is made up of the mean molecular-velocity V downwards and the velocity v parallel to AB. This latter velocity is the mass-velocity of the gas as distinct from its molecular velocity. Since all the molecules passing across AB either from above or from below are travelling with a velocity V normal to the plane, there is no change of momentum on this account. [The mass-velocity is small compared with V, so that the effect of v on the direction of the molecules crossing AB is neglected.]

Fig. 4.



The transfer of momentum in the y-direction per second, caused by the molecules crossing AB from above, is therefore $\frac{1}{6}\rho V\left(v+\lambda\frac{\partial v}{\partial z}\right)$, if we accept Joule's hypothesis that the mass of gas moving per second across the unit-plane in a direction normal to it is $\frac{1}{6}\rho V$, where ρ is the density of the gas. The transfer per second, due to those molecules crossing from a distance λ below AB, is $\frac{1}{6}\rho V\left(v-\lambda\frac{\partial v}{\partial z}\right)$. The aggregate transfer of momentum per second is therefore $\frac{1}{3}\rho\lambda V\frac{\partial v}{\delta z}$.

Similarly, the rate of transfer of momentum in the y-direction across a unit-plane near to AB but at a height $z+\delta z$ is $\frac{1}{3}\rho\lambda V\left[\frac{\partial v}{\partial z}+\frac{\partial}{\partial z}\left(\frac{\partial v}{\partial z}\right)\delta z\right]$. The nett rate of transfer of momentum in the y-direction to the gas in the layer between the unit-planes is therefore $\frac{1}{3}\rho\lambda V\frac{\partial^2 v}{\partial z^2}\delta z$. Thus the force on this layer of gas is $\frac{1}{3}\rho\lambda V\frac{\partial^2 v}{\partial z^2}.\delta z$. Another expression for this force is obtained by considering the forces acting on the gas in the above layer due to viscous forces in the gas. On the lower side the force is $\eta\frac{\partial v}{\partial z}$, and acts in the negative direction of y: on the upper side it is $\eta\left[\frac{\partial v}{\partial z}+\frac{\partial}{\partial z}\left(\frac{\partial v}{\partial z}\right)\delta z\right]$, and acts in the positive direction of y. Now although the lines of action of these forces are not concurrent, we may subtract them in order to find the force urging the gas forward and equate this to the force arising from the difference in pressure across the ends of the element, since the gas is not being accelerated. For our present purpose, however, we only require the force acting on the gas due to its viscosity. It is

$$\eta\frac{\partial^2 v}{\partial z^2}.\delta z.$$

Equating the two expressions for the forces thus obtained, we have

$$\eta=\tfrac{1}{3}\rho\lambda V.$$

Now the defect in the usual text-book proof is that the mass-velocity v is assumed, quite unnecessarily, to vary directly as the height z of the plane considered above xOy. To find the reason why a proof based on such an argument is invalid it is necessary to consider all the external forces acting on an element of area $\delta x\,\delta y$ and thickness δz. If p is the pressure at the end $y=$const., and $p+\frac{\partial p}{\partial y}\delta y$ that at the end $y+\delta y=$const., we have

$$-\frac{\partial p}{\partial y}=\eta\frac{\partial^2 v}{\partial z^2}.$$

When v is directly proportional to z, $\frac{\partial^2 v}{\partial z^2}$ is zero, so that the drop in pressure per unit length in the direction of the flow is zero: in other words, there is no mass-motion of the gas. It is hardly necessary to add that it is fallacious to obtain an expression for the viscosity of a gas in which the proof tacitly assumes that the gas is at rest.

The Conduction of Heat in Gases.

Let us suppose that AB, fig. 4 (*b*) is a small area $\delta x\,.\,\delta y$ in a gas where the temperature is θ. The height of AB above a reference plane xOy is z; at height $z+\delta z$ the temperature will be $\theta+\delta\theta$. In this problem there is no mass-velocity to be assigned to the gas, but it is assumed that the mass of gas passing downwards through AB is $\frac{1}{6}\rho V$ per second per unit area and that the average temperature of this gas is that of molecules in the plane at a height λ above AB, *i. e.*, its temperature is $\theta+\lambda\frac{\partial\theta}{\partial z}$. In the same time an equal amount of gas per unit area of AB will pass upwards, but at a temperature $\theta-\lambda\frac{\partial\theta}{\partial z}$. The nett transfer of heat per second across unit area of AB in the positive direction of z is therefore $-\frac{1}{3}\rho\lambda Vc_v\frac{\partial\theta}{\partial z}$, where c_v is the specific heat of the gas at constant volume. If κ is the thermal conductivity of the gas, we have

$$Q=-\kappa\frac{\partial\theta}{\partial z},$$

where Q is the heat passing per second across unit area of AB. Hence

$$\kappa=\tfrac{1}{3}\rho\lambda Vc_v=\eta c_v.$$

Osborne Reynolds' Apparatus for Investigating how the Viscosity of a Liquid varies with Temperature (Concentration).

The original apparatus was first designed in 1886, and concerning it Reynolds writes * "*Since the value for η for water is known for all moderate temperatures, in order to obtain the value for oil it is only necessary to ascertain the relative times taken by the same volumes of oil and water to flow through the same channel, care being taken to make the channel such that there are no eddies and that the energy of motion is small compared with the loss of head. These times are proportional to $\frac{\eta}{p}$, where p is the fall of pressure.*" As ordinarily carried out only a somewhat clumsy attempt is made to keep the head of liquid constant. The purpose of this note is to show how this difficulty may be overcome and also to set out the theory in such a way that no confusion about signs arises. In its usual form the apparatus consists of a glass siphon connected to a length of capillary tube, BC, fig. 5 (*a*), by means of a short piece of rubber tubing. The capillary tube is in a vertical position, and the glass-rubber joints must be "watertight." If only comparative measurements are to be made then the glass tubes may be fused together: if this is done the effective length of the capillary becomes more difficult to determine but eddies are less likely to be formed. P is a "glass pin," bent as shown, and the level, A, of the liquid in the beaker is adjusted so that the point of the pin coincides with it. The advantage of using the pin in this way is that the point of the pin can readily be adjusted to coincide with its image formed by reflexion in the liquid surface. The siphon is filled with the liquid under investigation and the mass of liquid exuding in a given time ascertained. From this data, and the dimensions of the apparatus, the value for η, the coefficient of viscosity of the liquid at the temperature of the experiment, may be calculated by means of the formula which we proceed to establish.

---

* Scientific Papers, vol. ii. p. 238 [or, Phil. Trans. clxxvii. pt. I., p. 157 (1886)].

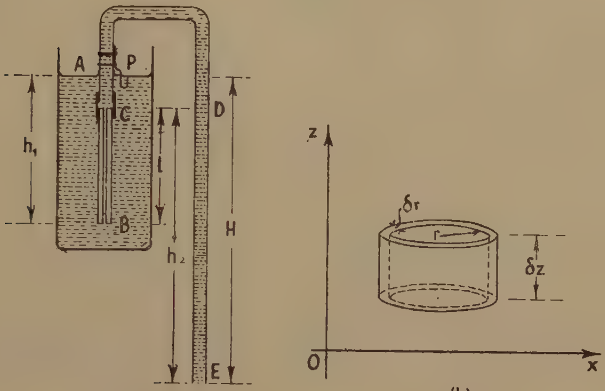
Let P_0 be the atmospheric pressure. Then pressure at $B=P_0+g\rho h_1$, where $h_1=AB$, g is the intensity of gravity, and ρ is the density of the liquid at the temperature of the experiment. The pressure at $E=P_0$; that at D=that at $C=P_0-g\rho h_2$, where $h_2=DE$.

$$\therefore \qquad P_B-P_C=g\rho(h_1+h_2).$$

Let z be the distance of a point in the capillary tube from B, the positive direction of z being upwards. The pressure gradient along the tube is constant, if the usual conditions of viscous flow are fulfilled. Hence

$$\frac{\partial p}{\partial z}=\alpha, \quad \text{or} \quad p=\alpha z+\beta,$$

Fig. 5.



(a) (b)

where α and β are constants to be determined from the end conditions. Let l be the length of the tube. Then

$$\text{at } z=0, \quad p=P_0+g\rho h_1, \quad \text{so that } \beta=P_0+g\rho h_1,$$

$$\text{and at } z=l, \quad p=P_0-g\rho h_2, \quad \text{so that } \alpha=-\frac{g\rho(h_1+h_2)}{l}.$$

$$\therefore \qquad \frac{\partial p}{\partial z}=-g\rho\,\frac{(h_1+h_2)}{l}.$$

This fixes the sign of the pressure gradient.

To determine the volume of liquid passing per second through the tube, consider a small cylindrical element with internal and external radii r and $(r+\delta r)$ respectively and length δz—*cf.* fig. 5 (*b*). Let p be the pressure on the lower end of the element. Then the force due to the pressure of the liquid on the lower plane face of the element is

$$2\pi r\,.\,\delta r\,.\,p$$

and acts upwards. On the upper plane face the force is

$$2\pi r\,.\,\delta r\left[p+\frac{\partial p}{\partial z}\,.\,\delta z\right],$$

and this acts downwards.

Resultant force due to pressure in the direction of z increasing is

$$-2\pi r \,.\, \frac{\partial p}{\partial z} \,.\, \delta r \,.\, \delta z.$$

The weight of the cylindrical element is $2\pi r\rho g \,.\, \delta r \,.\, \delta z$, and this acts vertically downwards.

If u and $u+\delta u$ are the velocities of flow at points r and $(r+\delta r)$ from the axis of the element, then Newton's definition of viscosity gives

$$2\pi r \,.\, \delta z \,.\, \eta\frac{\partial u}{\partial r} = 2\pi\eta \,.\, \delta z \,.\, r\frac{\partial u}{\partial r}$$

as the force due to viscosity acting on the inner curved surface, and this acts downwards. [The important point to emphasize here is that since $\frac{\partial u}{\partial r}$ is negative the force on the inner curved surface will actually act upwards, but no minus sign must be introduced into the above expression for the force, as is so often done, if the recognized practice and methods of the calculus are not to be violated.] On the outer curved surface the force due to viscosity is

$$2\pi\eta \,.\, \delta z\left[r\frac{\partial u}{\partial r}+\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right)\delta r\right],$$

and this acts upwards. Since the liquid is moving without acceleration

$$-2\pi r \,.\, \delta r \,.\, \delta z \,.\, \frac{\partial p}{\partial z} = 2\pi r \,.\, \delta r \,.\, \delta z\, \rho g - 2\pi\eta \,.\, \delta z \,.\, \delta r\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right),$$

i. e.,

$$\eta\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) = r\rho g + r\frac{\partial p}{\partial z} = r\left(g\rho+\frac{\partial p}{\partial z}\right).$$

Since $\frac{\partial p}{\partial z}$ is independent of r, integrating, we have

$$\eta\left[r\frac{\partial u}{\partial r}\right] = \frac{r^2}{2}\left[\rho g+\frac{\partial p}{\partial z}\right]+\mathrm{A},$$

where A is an integration constant.

$$\therefore \quad \eta\frac{\partial u}{\partial r} = \frac{r}{2}\left[g\rho+\frac{\partial p}{\partial z}\right]+\frac{\mathrm{A}}{r}.$$

Integrating again

$$\eta[u] = \frac{r^2}{4}\left(g\rho+\frac{\partial p}{\partial z}\right)+\mathrm{A}\log r+\mathrm{B},$$

where B is another integration constant. Now $\mathrm{A}=0$, since u is finite when $r=0$, and at $r=a$, $u=0$.

$$\therefore \quad \mathrm{B} = -\frac{a^2}{4}\left(g\rho+\frac{\partial p}{\partial z}\right).$$

and

$$\therefore \quad u = -\frac{1}{4\eta}\left(g\rho+\frac{\partial p}{\partial z}\right)(a^2-r^2).$$

Let Q be the volume of liquid flowing per second. Then

$$Q=-\int_0^a 2\pi r\,.\,dr\,.\,\frac{1}{4\eta}\left(g\rho+\frac{\partial p}{\partial z}\right)(a^2-r^2)=\frac{\pi}{2\eta}\left(g\rho+\frac{\partial p}{\partial z}\right)\left(-\frac{a^4}{4}\right)$$
$$=\frac{\pi a^4}{8\eta l}g\rho\,\mathrm{H},$$

where H is the depth of E below A.

$$\therefore \qquad Q=\frac{\pi a^4 \mathrm{P}}{8\eta l}\text{, where } \mathrm{P}=g\rho\mathrm{H}.$$

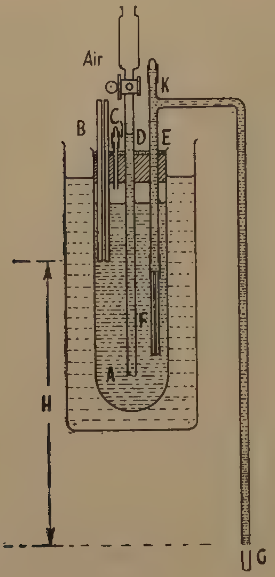
As the liquid runs out from the siphon the capillary tube should be lowered so that the point of the pin is always in contact with the liquid surface. At best, this method only maintains P approximately constant. If this apparatus is used to investigate how the viscosity of a liquid varies with temperature, then the temperature of the liquid in the siphon is not equal to that of the liquid as it flows through the capillary tube, and it is difficult to estimate the correction with any degree of accuracy. When investigating how the viscosity at room temperature of a solution varies with the concentration of the dissolved substance this difficulty does not arise.

A modified form of Osborne Reynolds' apparatus is shown in fig. 7 (*b*). It is true that the first difficulty just mentioned is not overcome, but the pressure head can be maintained constant when experiments at room temperature are performed. A large tube A is fitted with a rubber bung through which pass four tubes. The tube B is a capillary tube, about 1 mm. in diameter, its lower end reaching half-way down the tube A. C is a short tube which may be opened or closed, with respect to the atmosphere, by withdrawing the solid glass rod which is attached to its upper end by means of rubber tubing. D is a wider tube attached to a small reservoir so that liquid may be introduced into A; by passing a current of air through this tube the liquid in A can be well stirred before commencing an experiment. This is very necessary since the viscosity of most liquids changes very rapidly with temperature, and hence the temperature throughout A must be kept uniform and constant. E is a siphon to the end of which is attached, either by means of rubber tubing or by fusing the glass together, the capillary tube, F, through which the liquid flows. It is about 10 cm. long and has an internal diameter of about 0·02 cm. The tube K, attached to the siphon, enables the latter to be filled prior to beginning work: it is then closed and serves to collect any small air bubbles which may appear during the course of the experiment. If these bubbles are allowed to collect in the siphon the steadiness of the flow of liquid may be seriously affected.

When the temperature of the liquid in A has become steady (the thermometer is not shown) the plug is inserted at C and the stopcock in D closed. When the plug G at the lower end of the siphon is withdrawn, the liquid will exude from the apparatus: it will be replaced by air which enters through B, and the head of liquid remains constant providing the level of the liquid in A never falls below the lower end of B. The coefficient of viscosity is then calculated in the usual way.

It would of course be possible to arrange for the siphon to be almost entirely within A by providing this tube with a cork at its lower end,

Fig. 6.

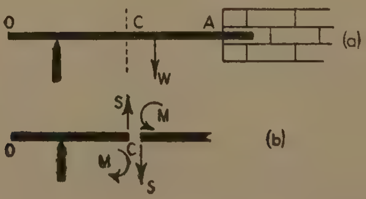


but the difficulties of maintaining a long column of a liquid at a constant temperature are so well known that it is difficult to believe that this additional modification would be beneficial.

On Bending Moments and the Bending of Beams.

The question of the sag in a horizontal beam is of paramount importance to engineers, and yet the subject of bending moments is generally presented in a poor manner. The main defect lies in the fact that due

Fig. 7.

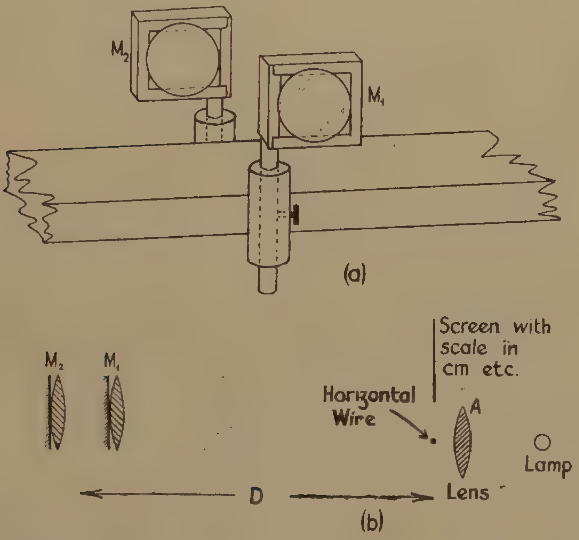


attention is not paid to signs. Suppose we have a beam supported and loaded in any manner so long as the deflexion produced is small. Such a beam is shown in fig. 7 (*a*). It is required to define the shearing force, S, and the bending moment, M, at the section C. Suppose that they are as indicated in fig. 7 (*b*). The origin of co-ordinates is O. Then, with this convention, when we state that the shearing force, S, in a given instance, is (say), 10 lb.-wt., we mean that the shearing force on the L.H. portion of the beam at this section is of magnitude 10 lb.-wt. and that it acts vertically upwards; on the R.H. portion the shearing force has the same magnitude but it acts downwards. Similarly, the bending moment, M, is positive when it acts in a clockwise direction on the L.H. portion of the beam.

Experimental Determination of Poisson's Ratio from Observations on a Uniformly Bent Beam.

When a light beam is supported symmetrically on knife-edges in a horizontal plane and loaded equally at its ends, it is well known that the portion of the beam between the knife-edges is bent into a circular arc, *i. e.*, the bending moment is constant at all sections in this portion of the beam and the shearing force is zero. It is easily shown, however, that the end portions are not uniformly bent; reference to this fact is generally omitted.

Fig. 8.



Moreover, it is well known that σ, Poisson's ratio for the material of the bar, is equal to the ratio

$$\frac{\text{Radius of curvature of the central portion}}{\text{Radius of anticlastic curvature of any section of the central portion}}.$$

In 1908 Searle devised a method for measuring the radius of anticlastic curvature. For several years the author has used the following modification of this method: it is more accurate:—An iron bar (80 cm. × 4 cm. × 0·33 cm.) is used. As shown in fig. 8 (*a*), two brass connectors are soldered to the bar at its central cross-section : these serve as small clamps. Two small plane mirrors, M_1 and M_2, are supported in copper frames soldered to short lengths of copper wire which fit the clamps. These copper wires may be slightly bent so that the mirrors are parallel to each other. Immediately in front of each mirror is fixed a converging lens of two metres focal length : soft wax holds each lens in position. A horizontal beam of light, from a filament lamp, is then directed on to the mirrors M_1 and M_2, fig. 8 (*b*), by means of a converging lens, A, with a fine horizontal wire stretched across it. The image of the filament is focussed on the mirrors. Two images of the above wire, one from each mirror and its attached lens, are obtained on a vertical scale graduated in cm., etc. on a translucent screen near to the lens A. This is possible on account of the large depth of focus possessed by a long-focus lens. When the beam is loaded symmetrically the images of the wire are displaced. If the mean distance, D, of the mirrors from the scale is measured, ϕ, the change in relative inclination of the mirrors can be deduced. When the width, b, of the beam is known, the radius of anticlastic curvature may be obtained at once. If d is the sum of the displacements of the images on the screen

$$\phi=\frac{d}{2\mathrm{D}}.$$

$$\mathrm{R}_2=\text{radius of anticlastic curvature}=\frac{b}{\phi}=\frac{2b\mathrm{D}}{d}.$$

If y_m is the numerical value of the deflexion at the middle of the beam, and R_1 the radius of curvature in the plane of bending, $2\mathrm{R}_1 \,.\, y_m=\frac{l^2}{4}$, where l is the distance between the knife-edges.

$$\therefore \qquad \frac{16\sigma b\mathrm{D}}{l^2}=\frac{d}{y_m}.$$

If, therefore, we plot $y_m=\mathrm{Y}$, and $d=\mathrm{X}$, a straight line with a slope $\frac{l^2}{16\sigma b\mathrm{D}}$ is obtained. Hence σ may be deduced.

On Clapeyron's Theorem of the Three Moments.

The proofs of this important theorem are often unsound on account of signs which are confused. Let ABC, fig. 9, be the points at which a uniformly loaded beam rests in a horizontal position upon three consecutive piers, a and b being the lengths of the two spans. Let $\mathrm{M_A}$, $\mathrm{M_B}$, and $\mathrm{M_C}$ be the bending moments at the points of support, while w is the weight of the beam per unit length. Take A as the origin of co-ordinates —x is measured along the bar ; y vertically downwards.

Then at any point in the first span, whose distance from A is x, i. e., $0<x<a$,

$$\mathrm{E}\mathrm{I}y''=\tfrac{1}{2}wx^2+\mathrm{A}_1x+\mathrm{B}_1,$$

where A_1 and B_1 are integration constants. When $x=0$, the bending moment is M_A, so that

$$M_A=B_1.$$

$$\therefore\quad EIy''=\tfrac{1}{2}wx^2+A_1x+M_A.$$

When $x=a$, the bending moment is M_B, so that

$$M_B=\tfrac{1}{2}wa^2+A_1a+M_A.$$

$$\therefore\quad EIy'=\tfrac{1}{6}wx^3+\tfrac{1}{2}A_1x^2+M_Ax+K,$$

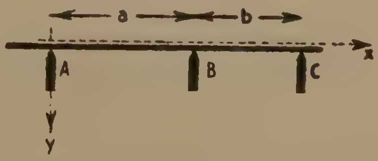
where K is an integration constant. Let the slope of the beam at $x=a$ be α, i. .e, $\left[\frac{dy}{dx}\right]_{x=a}=\alpha$. Then

$$EI\alpha=EI[y']_{x=a}=\tfrac{1}{6}wa^3+\tfrac{1}{2}A_1a^2+M_Aa+K.$$

$$\therefore\ EIy=\frac{1}{24}wx^4+\frac{1}{6}A_1x^3+\frac{1}{2}M_Ax^2+EI\alpha x-\left(\frac{1}{6}wa^3+\frac{1}{2}A_1a^2+M_Aa\right)x+J,$$

where J is a constant of integration.

Fig. 9.



Now when $x=0$, $y=0$, so that $J=0$. Also, when $x=a$, $y=0$. This gives A_1.

$$\therefore\quad M_B=\frac{1}{8}wa^2-\frac{3}{2}M_A+\frac{3EI}{a}\alpha+M_A,$$

or

$$aM_A+2aM_B-\tfrac{1}{4}wa^3=6EI\alpha.$$

Now consider the portion BC of the beam. To get our signs correct when we attempt to write down the corresponding equation to that just obtained, imagine that the whole beam has been turned round a vertical axis so that it is viewed from the other side. Then the angle which the beam makes with the horizontal at B is $(\pi-\tan.^{-1}\alpha)$, so that the slope is $-\alpha$. Hence

$$bM_C+2bM_B-\tfrac{1}{4}wb^3=-6EI\alpha.$$

$$\therefore\quad aM_A+2(a+b)M_B+bM_C=\tfrac{1}{4}w(a^3+b^3).$$

Thus the theorem is established.

The Cantilever.

(*a*) Suppose that the mass per unit length of a horizontal cantilever is negligible and that it is supported in such a manner that it can only be deflected in a *horizontal* plane. To deflect the lever a light string is attached to the end of the lever and after passing over a light pulley

(5 cm. in diameter at least) carries a scale pan : when the pan is loaded the beam is deflected and the deflexion can be measured both with the load in the pan increasing and also decreasing. Moreover, the force can be applied in the opposite direction, and thus a reliable mean value for the change in deflexion per unit change in the deflecting force may be obtained.

A suitable lever for experimental purposes consists of a wooden lath about 150–200 cm. long, 0·5 cm. thick and 2·5 cm. wide. The mass per unit length of such a lever is certainly not negligible, but if the lever is supported in the manner indicated the weight of the lever does not affect the deflexion.

If l is the length of the cantilever the deflexion at the end is given by

$$\zeta=[y]_{x=l}=\frac{\mathrm{W}}{\mathrm{EI}}\cdot\frac{l^3}{3}=\frac{mg}{\mathrm{EI}}\cdot\frac{l^3}{3},$$

where m is the load in the pan and g the intensity of gravity. To correct for the presence of any small frictional forces at the pulley, which are not eliminated by taking observations on each side of the equilibrium position of the cantilever, we may write

$$\zeta=\frac{(m+\alpha)g}{\mathrm{EI}}\cdot\frac{l^3}{3}$$

and determine both E and α graphically by plotting ζ against m.

(*b*) The above arrangement also permits one to determine the period of vibration of the lever when different masses are attached to its end. The motion is relatively slow, so that the transits of the end of the lever past a fiducial mark can be seen so that no stroboscopic method is necessary to find the periodic time T. The mass of the lever, however, cannot be neglected.

Let W be the deflecting force (in a horizontal plane). Then the deflexion y at A, fig. 10 (a), at a distance x from the fixed end, is given by

$$\mathrm{EI}y=\tfrac{1}{2}\mathrm{W}lx^2-\tfrac{1}{6}\mathrm{W}x^3.$$

At $x=l$,

$$\mathrm{EI}\zeta=\tfrac{1}{3}\mathrm{W}l^3.$$

$$\therefore\quad y=\zeta\left[\frac{3}{2}\frac{x^2}{l^2}-\frac{1}{2}\frac{x^3}{l^3}\right].$$

If μ is the mass per unit length of the lever, its kinetic energy is

$$\int_0^l\frac{1}{2}\mu\left(\frac{3}{2}\frac{x^2}{l^2}-\frac{1}{2}\frac{x^3}{l^3}\right)^2\dot{\zeta}^2dx=\frac{33}{280}\mu l\dot{\zeta}^2.$$

The kinetic energy of the mass M carried at the end of the lever is $\frac{1}{2}\mathrm{M}\dot{\zeta}^2$.

The potential energy of the beam is equal to the work done in straining it, *i. e.*,

$$\int_0^\zeta\frac{3\mathrm{EI}}{l^3}\zeta\,.\,d\zeta=\frac{3}{2}\frac{\mathrm{EI}}{l^3}\cdot\zeta^2.$$

Since the total kinetic energy+the potential energy is constant,

$$\left(\frac{1}{2}\mathrm{M}+\frac{33}{280}\mu l\right)\dot{\zeta}^2+\frac{3}{2}\frac{\mathrm{EI}}{l^3}\zeta^2=\text{constant}.$$

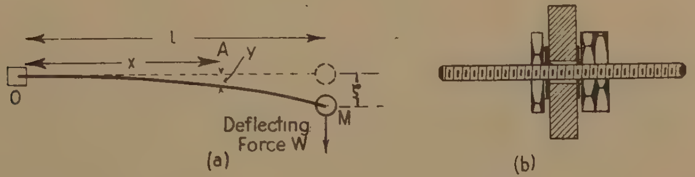
$$\therefore \quad \left(\frac{1}{2}\mathrm{M}+\frac{33}{280}\mu l\right)\ddot{\zeta}+\frac{3}{2}\frac{\mathrm{EI}}{l^3}\zeta=0.$$

$$\therefore \quad \mathrm{T}=2\pi\sqrt{\frac{\left(\mathrm{M}+\frac{33}{140}\mu l\right)l^3}{3\mathrm{EI}}}.$$

Again, a graphical method, in which T^2 is plotted against M, serves to determine both E and μ. The validity of the above theory can therefore be checked.

To attach the masses M to the lever a brass screw of known mass is passed through the beam and fixed in position by brass nuts of known mass. Additional " nuts " can be added to vary the mass—*cf.* fig. 10 (*b*).

Fig. 10.



If such a lever is supported so that it can execute vibrations in a *vertical plane*, the curvature is considerable even when the beam is unloaded so that the simple theory of bending is violated : of course, if short levers are used this question does not arise, but more elaborate methods have to be used to time the motion.

Specific Heat and Thermal Capacity.

If m is the mass of a body and θ the rise in temperature when a quantity of heat Q is added to it, we have

$$\mathrm{Q} \propto m\theta,$$

and therefore write

$$\mathrm{Q}=sm\theta,$$

where s is a constant known as the specific heat of the material of the body. This definition shows at once that in the C.G.S. system the dimensions of s are [cal. gm.$^{-1}$ deg.$^{-1}$]. The quantity ms is termed the thermal capacity of the body.

[In more advanced thermodynamics it would of course be better to write

$$\delta\mathrm{Q}=sm\delta\theta,$$

where s is the specific heat of the substance at a temperature θ.]

The confusion which still exists is shown by the following remarks taken from a recent paper by Hansel *. This author says "*I favour the definition of specific heat given in Chambers' Dictionary :—'Specific heat, the number of heat-units necessary to raise the unit of mass of a given substance one degree in temperature.' This definition is dimensionally correct and applies to all consistent sets of units.*"

The definition is only correct numerically, certainly not dimensionally. In an elementary course in physics one might be permitted to say that the specific heat of a substance is numerically equal to the amount of heat necessary to raise the temperature of unit mass by one degree. Later, in his paper, Hansel suggests a definition of specific heat which agrees with that given on p. 791 †.

Specific (Intrinsic) Entropy and Entropy.

In thermodynamics little attention is paid nowadays to the difference between these two quantities. If δq is the heat added (mechanical units) in a reversible process to unit mass of a substance at absolute temperature T, then $\frac{\delta q}{T}$ is the change in the specific entropy and is denoted by δS, *i. e.*,

$$\delta S = \frac{\delta q}{T},$$

and the units for specific entropy will be, in the C.G.S. system, erg. gm.$^{-1}$ deg.$^{-1}$K. [or cal. gm.$^{-1}$ deg.$^{-1}$K. if thermal units are used].

Now the mass of a substance times its change in specific entropy is the change in entropy of the body and is denoted by $\delta\phi$, *i. e.*,

$$\delta\phi = m \cdot \delta S,$$

where m is the mass of the body. The dimensions of $\delta\phi$ are accordingly erg. deg.$^{-1}$K. [or cal. deg.$^{-1}$K.].

The importance of making this distinction between the variables S and ϕ is very apparent when the dimensions of thermodynamical equations are considered. Thus, the change in specific entropy of a liquid as it passes completely into a saturated vapour at the same absolute temperature T is given by

$$S_2 - S_1 = \frac{l}{T},$$

where l is the latent heat of evaporation of the liquid at temperature T. Now $\frac{l}{T}$ has dimensions $\left\{\begin{matrix}\text{erg.}\\ \text{cal.}\end{matrix}\right\}$ gm.$^{-1}$ deg.$^{-1}$ K., which are those of specific entropy and not of entropy.

The confusion arising from the fact that the distinction between entropy and specific entropy is not emphasized is further illustrated in the ensuing paragraph.

---

* Hansel, Proc. Phys. Soc. liv. p. 159 (1942).

† *Cf.* C. J. Smith, 'Intermediate Physics' (1935).

Surface Energy and Surface Tension.

Let a given surface of area σ expand by a small amount $\delta\sigma$, and let δT be the corresponding increase in temperature. Let E be the surface energy of the whole surface—not to be confused with ϵ, the surface energy per unit area. Then E will be a function of T and σ. Thus

$$dE=\left(\frac{\partial E}{\partial T}\right)_\sigma dT+\left(\frac{\partial E}{\partial \sigma}\right)_T d\sigma$$

$$=c\,dT+\epsilon\, d\sigma, \quad \ldots \ldots \ldots \quad (1)$$

where c is written for $\left(\frac{\partial E}{\partial T}\right)_\sigma$ and $\left(\frac{\partial E}{\partial \sigma}\right)_T$ is the surface energy per unit area at temperature T, which has been designated ϵ.

The heat which must be supplied to accomplish this change is given by

$$dQ=dE+dW=dE-\gamma\, d\sigma,$$

where γ is the surface tension. [N.B.—We do not write dq, which would refer to unit area of the surface.] The corresponding change in entropy of the surface is given by

$$d\phi=\frac{dQ}{T}=\frac{c}{T}dT+\frac{\epsilon-\gamma}{T}d\sigma.$$

It must be noted that this equation expresses the change in entropy of the surface and not the change in specific entropy which, in this instance, would be the change in entropy per unit area.

Since $d\phi$ is a perfect differential, we have

$$\left[\frac{\partial}{\partial\sigma}\left(\frac{c}{T}\right)\right]_T=\left[\frac{\partial}{\partial T}\left(\frac{\epsilon-\gamma}{T}\right)\right]_\sigma . \quad \ldots \ldots \quad (2)$$

But $\left(\frac{\partial c}{\partial \sigma}\right)_T=\left(\frac{\partial \epsilon}{\partial T}\right)_\sigma$. [$\because$ dE is a perfect differential.]

Hence $$\left(\frac{\partial \epsilon}{\partial T}\right)_\sigma=T\left[\frac{\partial}{\partial T}\left(\frac{\epsilon-\gamma}{T}\right)\right]_\sigma . \quad \ldots \quad (3)$$

$$\therefore \qquad \epsilon=\gamma-T\left(\frac{\partial\gamma}{\partial T}\right)_\sigma,$$

from which ϵ may be deduced when γ and $\left(\frac{\partial\gamma}{\partial T}\right)_\sigma$ are known.

On Drawing Tangents to Curves: Experimental Determination of the Specific Heat of a Liquid by the Method of Cooling.

In a recent paper, read before a joint meeting of the Science Master's Association and the Physical Society, it was stated "*It is impossible to obtain rates of cooling with sufficient accuracy by drawing tangents to a cooling curve*[*]." For several years the author has used the following method and has recently embodied in it the device described by Calthrop[†].

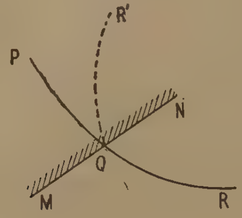
---

[*] Hansel, Proc. Phys. Soc. liv. p. 163 (1942).

[†] 'Advanced Experiments in Practical Physics,' p. xv.

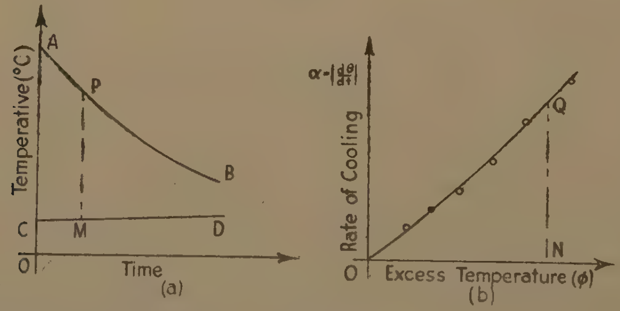
Suppose it is required to draw a tangent at a point Q on a curve PQR (fig. 11). A small plane mirror MN, with its plane normal to that of the diagram, is placed across the curve at Q as indicated. QR′ is the image of the portion QR of the curve, so that the curve appears to be broken at Q ; by rotating the mirror about a vertical axis through R, an image of RQ, which is continuous with RQ, is obtained. The mirror is then normal to the curve at Q, so that the tangent is readily constructed.

Fig. 11.



To obtain a still better value for the slope of the curve at a specified point, let us assume that a determination of the specific heat of a liquid is being made by the well-known "method of cooling." In fig. 12 (*a*), AB is the curve showing how the actual temperature θ of the calorimeter and its contents varies with time, and CD shows the march of the room temperature in the same interval. Suppose that we wish to find the slope of the curve at P when the excess temperature is PM, or ϕ. The normals and tangents at several points along AB are constructed, as described above, and the value of $\alpha = \left|\frac{d\theta}{dt}\right|$ at each is determined. The

Fig. 12.



(a)

(b)

"curve" shown in fig. 12 (*b*) is then constructed and if ON represents the excess temperature PM, by drawing the ordinate through N to cut the curve in Q we obtain QN, which is an excellent value for α appropriate to the point P on AB. The curve for the standard liquid (water) must be dealt with in the same way.

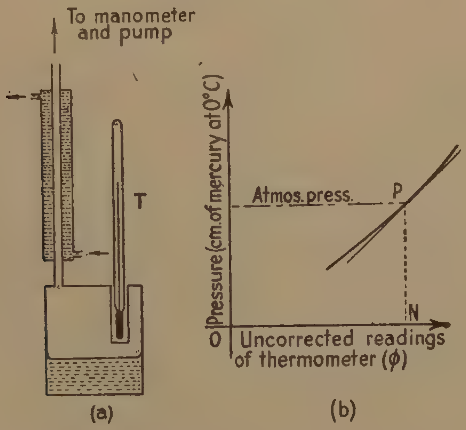
It is hardly necessary to add that this method of measuring the specific heat of a liquid is independent of the law of cooling and that it gives the specific heat of the liquid at a definite temperature. In actual practice the "curve" OQ is very nearly a straight line. To obtain accurate results by this method the calorimeter should be blackened externally in order to render more definite the main factor determining the rate at which heat is lost from the surface of the calorimeter: moreover, a lid to fit the calorimeter is essential to prevent loss of heat by evaporation.

This method of obtaining a good value for the slope of a curve at any point is not limited to the above experiment.

On Regnault's Method of Determining the Variation of the Saturation Vapour Pressure of Water with Temperature.

As a laboratory exercise students at this college have used the following apparatus in which some of the essential features of Regnault's apparatus are retained. The apparatus, fig. 13 (*a*), is constructed entirely of brass and the thermometer T is inserted into a brass tube arranged as shown.

Fig. 13.

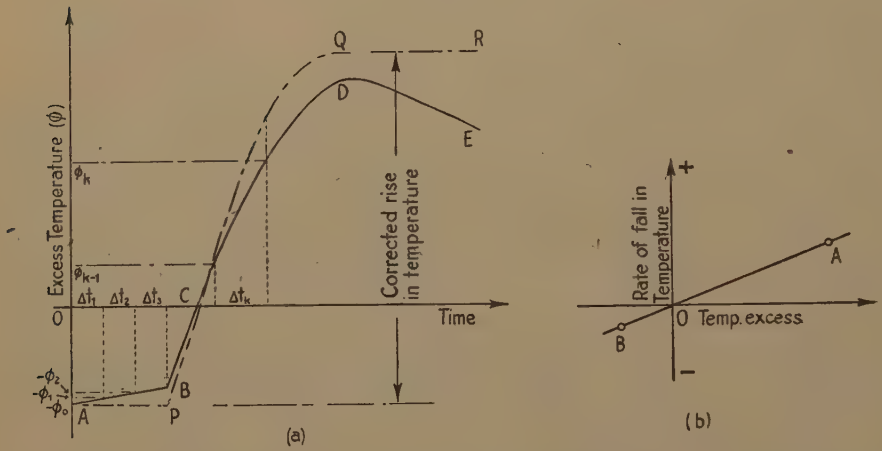

To manometer and pump
T
(a)
Pressure (cm. of mercury at 0°C)
Atmos. press.
P
N
O
Uncorrected readings of thermometer (ϕ)
(b)

The bulb of the thermometer is at the temperature of the vapour in the hypsometer and no error in its indication can be caused by mechanical deformation of the bulb due to variations in pressure on it. Suppose a set of observations has been made, preferably beginning with the pressure inside the apparatus at its lowest value. A graph showing how p (the saturation vapour pressure in cm. of mercury at 0° C.) varies with ϕ, the uncorrected reading of the thermometer, is constructed—*cf*. fig. 13 (*b*). Now to correct the thermometer readings we are not justified in assuming that a change in pressure of 1 cm. of mercury alters the normal boiling point of water by 0·37° C., for this result is actually obtained from the experiment just performed. By drawing the tangent at P, as described on p. 793, however, a value for $\frac{d\phi}{dp}$ which is an approximate value for $\frac{d\theta}{dp}$ is obtained [θ is the temperature on an ideal thermometer], and this can be used to calculate the correction to be applied to the mercury thermometer at the steam point. The readings of the thermometer can then be corrected in the usual way, a new graph plotted and, if necessary, a value for $\frac{d\theta}{dp}$ more nearly approaching the correct value obtained and the process repeated.

Regnault's Method for Determining the Correction for the Heat Exchange between a Calorimeter and its Surroundings in Calorimetric Experiments.

It is frequently stated that in calorimetric work a correction is necessary for the heat loss by radiation. The wording is erroneous for heat may be gained and, in general, heat will be lost or gained by conduction and convection. It is therefore desirable to speak about a correction for the heat exchange between a calorimeter and its surroundings.

Fig. 14.



(a) (b)

When Regnault's method for estimating this correction is explained to students it is often assumed that the initial temperature of the calorimeter is equal to that of the surroundings. This is undesirable since much benefit may be derived from carrying through the argument in a quite formal and general manner : moreover, if the initial and final temperatures are respectively below and above the temperature of the surroundings the correction is made smaller.

Suppose that ABCDE, fig. 14 (*a*), is the heating-curve obtained in the usual way : the ordinates denote excess temperature. Let time intervals Δt_1, Δt_2, . . . Δt_k, ... be marked off on the time axis and let the corresponding ordinates be drawn. In fig. 14 (*b*) let A and B denote the mean rates of *fall* in temperature in the intervals corresponding to the portions AB and DE of the heating-curve. Join OA and OB : these lines are not strictly collinear on account of the fact that the area of the surface of the calorimeter in contact with the liquid is different in the two cases, since the hot body has been introduced into the calorimeter when the portion of the curve DE is obtained. Thus one ought not to join A and B by a straight line. The following Table is then constructed:—

Time interval.	Excess temps. for this interval.		Mean excess temp.	Rate of *cooling* at mean excess temp.	Correction for heat exchange in interval.	Corrected temp. at end of interval.
	(Initial.)	(Final.)				
Δt_1	$-\phi_0$	$-\phi_1$	$-\frac{1}{2}(\phi_0+\phi_1)$	a_1 ($-$ve in such a case.)	$a_1 . \Delta t_1$	$-\phi_1+\alpha_1 . \Delta t_1$
Δt_2	$-\phi_1$	$-\phi_2$	$-\frac{1}{2}(\phi_1+\phi_2)$	a_2	$a_2 . \Delta t_2$	$-\phi_2+\sum_1^2 a_n . \Delta t_n$
. . .	. . .	. . .	. . .	. . .	. . .	. . .
Δt_k	ϕ_{k-1}	ϕ_k	$\frac{1}{2}(\phi_{k-1}+\phi_k)$	a_k ($+$ve in such a case.)	$\alpha_k . \Delta t_k$	$\phi_k+\sum_1^k a_n . \Delta t_n$
. . .	. . .	. . .	. . .	. . .	. . .	. . .

The corrected heating-curve APQR is then drawn: from it the rise in temperature of the calorimeter and its contents, corrected for heat exchange between the calorimeter and its surroundings, is obtained at once.

The justification for the procedure adopted above is as follows:—

In the time interval Δt_k we have

Heat lost in time $\Delta t_k = \Sigma(ms)$ [rate of cooling at mean temp.] . Δt_k.

$\therefore$ Heat received by calorimeter and contents in this interval

$$=\Sigma(ms)[\phi_k-\phi_{k-1}+\alpha_k . \Delta t_k],$$

i. e., by increasing the final temperature in the interval by $\alpha_k \Delta t_k$ the required corrected temperature is obtained.

Thermal Conductivity.

The author has for many years advocated the abolition of the "unit cube idea" wherever it is used in connexion with certain definitions in physics. The objection to the use of this cube is that when it is introduced the definition is only valid numerically for the dimensions are suppressed. For example, to define the thermal conductivity of a substance (generally in the form of a bar) one supposes that Q is the heat passed per second across an area A when the steady state has been reached and the direction of the heat flow is entirely normal to A. Then

$$\frac{Q}{A} \propto \text{fall in temperature per unit distance along the bar at the section considered},$$

i. e.,

$$\frac{Q}{A}=\kappa \text{ [fall in temp. . . . .]}$$
$$=-\kappa\frac{d\theta}{dx},$$

where θ is the temperature at a distance x, and $\theta+\delta\theta$ that at $x+\delta x$ along the bar, and where κ is a constant for the material of the bar [at the temperature considered]; κ is called the thermal conductivity of the material of the bar [at the temperature θ]. [With elementary students the calculus need not be introduced.] The advantage of this definition is that it gives the dimensions of κ at once, viz.,

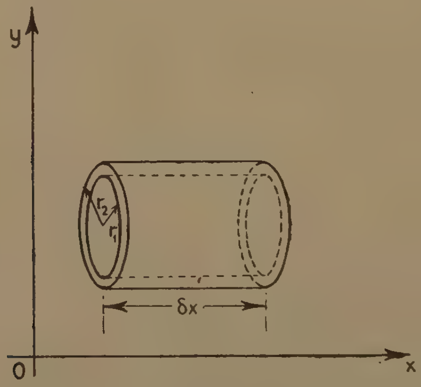
$$[\kappa]=\text{cal. cm.}^{-1}\text{ sec.}^{-1}\text{ deg.}^{-1}\text{ C.},$$

if the unit of heat is the calorie and C.G.S. units are used.

Experimental Determination of the Thermal Conductivity of a Material available in the Form of a Tube.

The theory of this laboratory experiment is often a stumbling block to students on account of the fact that sign conventions are violated in the proofs given to them.

Fig. 15.



Let r_1 and r_2, fig. 15, be the internal and external radii of the coaxial surfaces of the tube, which is surrounded by steam and through which a steady stream of cold water flows. Its length is defined by $x=0$, $x=l$. Let $x=0$ be the end at which cold water enters the tube and consider the rate of flow of heat across an element of the tube whose ends are defined by the planes x and $x+\delta x$ when steady conditions have been attained. Let θ_s be the steam temperature, ϕ the temperature of the water within this element and hence of the surface whose radius is r_1. Let θ and $\theta+\delta\theta$ be the temperatures of the surfaces defined by r and $r+\delta r$. Then δQ_0, the heat flowing *outwards* per second across the element is given by

$$\delta Q_0=-\kappa\,.\,2\pi r\,.\,\delta x\,.\frac{d\theta}{dr}.$$

But this heat flow is the same for every coaxial cylindrical element into which this length of the tube may be divided.

$$\therefore \quad \delta Q_0 \cdot \int_{r_1}^{r_2} \frac{dr}{2\pi r} = -\kappa \delta x \cdot \int_{\phi}^{\theta_s} d\theta.$$

$$\therefore \quad \delta Q_0 = \frac{-2\pi\kappa(\theta_s - \phi)\delta x}{\log_\varepsilon \frac{r_2}{r_1}}.$$

Thus the rate at which heat is flowing outwards across the whole length of the tube is given by

$$Q_0 = -\frac{2\pi\kappa}{\log_\varepsilon \frac{r_2}{r_1}} \int_0^l (\theta_s - \phi) dx.$$

To proceed further, it is assumed that ϕ is a linear function of x, *i. e.*, $\phi = \phi_1(1+\gamma x)$, where $\phi = \phi_1$ at $x=0$ and $\phi = \phi_2$ at $x = l$;

$$\therefore \quad \phi = \phi_1\left(1 + \frac{\phi_2 - \phi_1}{l} \cdot x\right).$$

Hence
$$Q_0 = -\frac{2\pi\kappa}{\log_\varepsilon \frac{r_2}{r_1}}\left[\theta_s - \frac{\phi_1 + \phi_2}{2}\right] l.$$

Thus, if m is the mass of water, its specific heat being taken as 1 cal. gm.$^{-1}$ deg.$^{-1}$ C., flowing per second, the heat flowing *inwards* through the walls of the tube per second is given by

$$Q = m(\phi_2 - \phi_1) = +\frac{2\pi\kappa}{\log_\varepsilon \frac{r_2}{r_1}}\left[\theta_s - \frac{\phi_1 + \phi_2}{2}\right] l,$$

since $Q = -Q_0$.

In practice one is never certain of the effective length of the tube and it is difficult to obtain a tube whose surfaces are coaxial. The author sought to overcome the first of these difficulties by using two thermocouples arranged at a fixed distance apart along the tube. At the same time, in order to test the linearity of ϕ, some six thermocouples were arranged at known intervals along the tube, the actual junctions lying close to the inner surface of the tube and being arranged spirally round its axis. The temperature distribution was very nearly linear but the method gave poor results because the temperatures were not measured actually at points on the inner glass surface. Present conditions prevent any further arrangements from being tested.

The Rate of Loss of Heat from a Full Radiator [*Black Body*].

If m is the mass of a black body at (absolute) temperature T, s the specific heat of its material, then the rate of loss of heat from it is given by

$$-ms\frac{dT}{dt} = \frac{\sigma A}{J}(T^4 - T_0^4),$$

where t is the time, σ Stefan's constant, A the surface area of the body, T_0 the temperature of the enclosure, and J is the mechanical equivalent of heat. In this equation it is tacitly assumed that T is the temperature throughout the body: this is only true if the thermal conductivity of its material is infinite. Moreover, this well-known equation is often given to students without any attempt being made to explain why the term in T_0 occurs in it. The body is certainly losing heat at a rate $\frac{\sigma A}{J}T^4$, if we accept the Stefan-Boltzmann law, but what is the justification for writing $\frac{\sigma A}{J}T_0^4$ as the rate at which the hot body is receiving heat when the area of the walls of the enclosure is certainly very different from that of the cooling body? The explanation is at once forthcoming if we consider the state of affairs when the body has itself reached a temperature T_0. Then it emits heat at a rate $\frac{\sigma A}{J}T_0^4$: but it is also receiving heat at this rate since its temperature remains constant. This must have been the rate at which it received heat from the walls of the enclosure when its temperature was T: our equation is therefore justified.

Planck's Law of Radiation: Wien's Displacement Law.

It is well known that all the fundamental laws of radiation are contained in Planck's law of radiation, viz.,

$$E_\lambda = \frac{8\pi hc}{\epsilon^{\frac{hc}{k\lambda T}} - 1},$$

where the symbols have their usual meanings. To determine the wavelength λ_m, corresponding to the maximum energy density, it is necessary to solve the equation

$$5(1-\epsilon^{-x})=x.$$

Planck* gives only one root to this equation, viz., $x=4\cdot9651$. Authors of text-books dealing with the subject of radiation are generally content to quote this solution and add remarks of which the following are typical.

> "*This is the only real root, as may be seen readily by drawing the curve.*"
> "*The root of this equation is* 4·96 *approximately.*"
> "*The solution to this equation is* $x=4\cdot9651$."
> "*This is a transcendental equation, but it can easily be seen that x is very nearly equal to* 5. *Actually it is* 4·9651."
> "*Equation dont la solution numérique est* $x=4\cdot9651$."

Each is misleading. Actually there are two real roots to the above equation, viz., $x_1=0$ and $x_2=4\cdot9651$. This fact was given by Richtmyer † in 1928, but the above quotations are from books published since then. It is remarkable how an error, once made, persists for several decades.

---

* Planck, *Ann. der Phys.* iv. p. 556 (1901).

† Richtmyer, 'Introduction to Modern Physics,' p. 242 (1928).

The solution $x_1=0$ corresponds to $\lambda\to\infty$ and so has no value in connexion with the physical problem in hand, but it should not be omitted from a general solution of the above equation.

Notation in Optics.

If μ_1 and μ_2 are the absolute refractive indices of two media for light of a certain frequency, θ_1 and θ_2, the angles of incidence and refraction respectively, then the second law of refraction, takes the form

$$\mu_1 \sin\theta_1=\mu_2 \sin\theta_2, \quad \text{or} \quad \Delta[\mu \sin\theta]=0.$$

Thus $\mu \sin\theta$ is an optical invariant.

The advantages of this formula are that by its aid problems in optics are greatly simplified, *e. g.*, if θ is the critical angle for rays of light travelling in a medium (1) and incident upon a medium (1)-medium (2) interface, then

$$\mu_1 \sin\theta=\mu_2,$$

and numerical exercises on the "critical angle" are solved at once.

Moreover, greater attention might be paid in the teaching of optics to obtaining equations which are symmetrical, *e. g.*, for the refraction of paraxial rays at a curved interface of radius r

$$\frac{\mu_2}{x_2}-\frac{\mu_1}{x_1}=\frac{\mu_2-\mu_1}{r},$$

or

$$\mu_2\left[\frac{1}{x_2}-\frac{1}{r}\right]=\mu_1\left[\frac{1}{x_1}-\frac{1}{r}\right],$$

where x_1 and x_2 are used instead of the more familiar "u" and "v." Then y_1 and y_2 can be used to denote the sizes of the object and image. Also, in the above instance, the linear magnification is given by

$$m=\frac{\dfrac{x_2}{\mu_2}}{\dfrac{x_1}{\mu_1}}.$$

By writing $\mu_1=-\mu_2$ in these and similar equations, equations valid for the reflexion of paraxial rays at curved surfaces are obtained.

If, as in dealing with the theory of coaxial refracting surfaces, it becomes necessary to refer to the position of an intermediate image, x' may be used to denote its position.

When distances of object and image are measured from the first and second principal foci respectively, it is now suggested that these distances be denoted by X_1 and X_2, so that Newton's equation becomes

$$X_1X_2=-f^2.$$

Similarly, for the case of images produced by the refraction of paraxial rays through a homogeneous sphere, it is more convenient to measure distances from the centre of the sphere. Denoting those of the object and image by ζ_1 and ζ_2 respectively, we have

$$\frac{1}{\zeta_2}-\frac{1}{\zeta_1}=\frac{2}{\rho}\left(\frac{\mu_1}{\mu_2}-1\right).$$

where ρ is the radius of the sphere considered numerically.

The Achromatism of Two Thin Lenses not in Contact.

The elementary discussion of this important problem is usually far from satisfactory. Suppose f_1 and f_2 are the focal lengths of two thin lenses separated by a distance a; let the powers of the lenses be κ_1 and κ_2. Then it is well known that

$$x_2=\frac{\alpha x_1+\beta}{\gamma x_1+\delta},$$

where α, β, γ, δ are constants for the particular system under consideration. Now the linear magnification, m, is given by

$$m=\alpha-\gamma x_2=(a\kappa_1+1)-(a\kappa_1\kappa_2+\kappa_1+\kappa_2)x_2.$$

$$\therefore \quad \delta(m)=a\omega_1\kappa_1-(a\omega_1\kappa_1\kappa_2+a\omega_2\kappa_1\kappa_2+\omega_1\kappa_1+\omega_2\kappa_2)x_2,$$

where the ω's are dispersive powers.

Hence, two coloured images will have the same size if

$$\delta(m)=a\omega_1\kappa_1-(a\omega_1\kappa_1\kappa_2+a\omega_2\kappa_1\kappa_2+\omega_1\kappa_1+\omega_2\kappa_2)x_2=0.$$

When this condition is satisfied the two coloured rays from one point in the object will emerge parallel to each other after refraction through the system on account of Helmholtz's equation $\Delta[\mu y\alpha]=0$.

The condition just obtained depends upon the position of the image for the equation involves x_2; it will only be independent of x_2 if a is zero or if $\omega_1=0$. The first condition violates the assumptions made in the problem and the second condition is impossible. for there is no material which will refract a multi-coloured beam of light without producing dispersion at the same time.

The condition for achromatism may, however, be fulfilled for a particular value of x_2. If this can be done the achromatism will be of a particular type and, even so, it must be remembered that it is partial in that only equality in the sizes of two coloured images will have been effected. Thus, if the system is to be achromatic, in the above restricted sense, when $x_2\rightarrow\infty$ then

$$a(\omega_1+\omega_2)\kappa_1\kappa_2+\omega_1\kappa_1+\omega_2\kappa_2=0,$$

and if the lenses are made of the same kind of glass, this equation reduces to

$$a=-\tfrac{1}{2}(f_1+f_2).$$

The treatment just given differs from the usual treatment as follows :— Ordinarily the expression used for m is

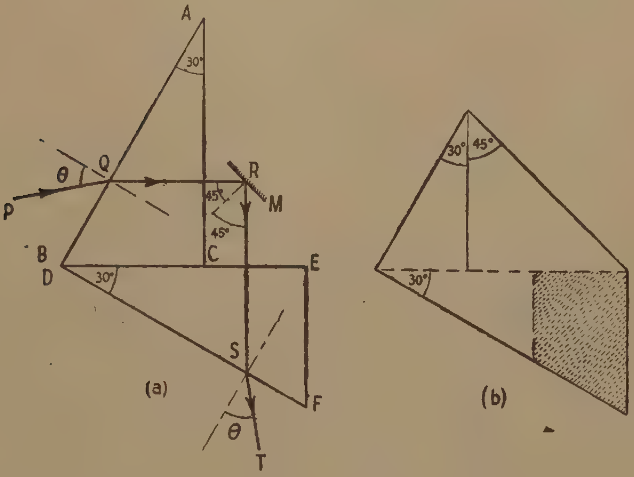
$$m=(\gamma x_1+\delta)^{-1},$$

and the system is achromatized in the above restricted sense for a particular position $(x_1 \rightarrow \infty)$ of the object. With eye-pieces this is impossible.

The Prism of the Constant Deviation Spectrometer.

One of the essential conditions for the production of a pure spectrum is that the rays shall pass symmetrically through the prism (60°): in practice this can only be done for light of one particular frequency. In dealing with the prism of a constant deviation spectrometer the fact that the above essential condition is not violated is seldom stressed. The principle of this prism is as follows :—Let ABC, fig. 16 (*a*), be one half of a 60° prism and suppose that PQ is a ray of light incident at such an angle θ that the refracted ray QR is parallel to BC. Let DEF be the other half of the 60° prism placed as shown. The ray QR emerges from the first half-prism without change of direction : assume that it falls on a plane mirror M so placed that the angle of incidence is 45°. The reflected ray will enter the prism DEF in a direction parallel to EF. The emergent ray will be ST and the angle of emergence θ. Thus the ray of light will pass through the compound arrangement of prisms and mirror just as if it had passed through the original 60° prism in the position of minimum deviation. The difference, however, is that the emergent ray ST is normal to the incident ray PQ.

Fig. 16.



If, instead of using a plane mirror, a 45° prism is placed as shown in fig. 16 (*b*), the same effect will be produced : and the portion of the glass prism shown shaded can be added so that the prism may be manufactured as an entity.

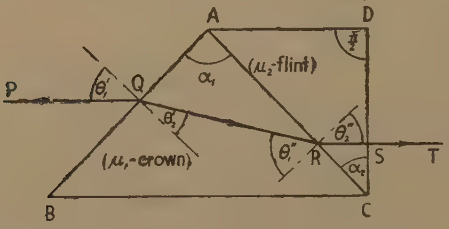
Amici's Prism and the Direct Vision Spectroscope.

In 1860 Amici * introduced a double prism of the form shown in fig. 17. ADC is a flint glass prism with a right-angle at D : it will be assumed that its refracting angle α_2 is known. ABC is a crown glass prism cemented to ADC along AC by means of Canada balsam. It is assumed that the indices of refraction μ_1 and μ_2 are known for the materials of the prisms ABC and ADC respectively for some standard ray—*e. g.* sodium light. The problem is to calculate the angle α_1 in order that a ray PQ, incident on AB in a direction parallel to AD, shall emerge parallel to itself, *i. e.*, along a normal to the face CD. The deviation for such a ray is zero. Let PQRST be the path of such a ray, the angles of incidence and refraction at different interfaces being as indicated. Then

$$\sin\theta_1'=\mu_1\sin\theta_2', \quad \ldots\ldots\ldots\ldots \quad (1)$$

$$\mu_1\sin\theta_1''=\mu_2\sin\theta_2'', \quad \ldots\ldots\ldots\ldots \quad (2)$$

Fig. 17.



and $\theta_2''=\alpha_2$, on account of the geometry of the system. Also

$$\theta_1'-\theta_2'=\theta_1''-\theta_2''. \quad \ldots\ldots\ldots\ldots \quad (3)$$

$$\therefore \quad \theta_1'=\alpha_1-\alpha_2.$$

Hence, from (1),

$$\sin(\alpha_1-\alpha_2)=\mu_1\sin(\alpha_1-\theta_1'').$$

$$\therefore \qquad \tan\alpha_1=\frac{\mu_1\sin\theta_1''-\sin\alpha_2}{\mu_1\cos\theta_1''-\cos\alpha_2}.$$

Eliminating θ_1'', we get

$$\tan\alpha_1=\frac{(\mu_2-1)\sin\alpha_2}{+\sqrt{\mu_1^2-\mu_2^2\sin^2\alpha_2}-\cos\alpha_2}.$$

When a beam of white light travelling in the direction PQ falls on an Amici prism all the rays except the sodium ray will be deviated, *i. e.*, dispersion will have been obtained. If the emergent beam falls on a converging lens each system of rays will be brought to a focus and a spectrum produced. Such an arrangement constitutes the most simple form of direct vision spectroscope. It suffers from the defect that the dispersion is small and that the sodium ray suffers a lateral displacement. If, however, two identical Amici prisms are arranged with their flint prisms adjacent, a direct vision spectroscope with greater dispersion will be obtained and, moreover, there will be no lateral displacement of the emergent ray. In practice only one flint prism, whose base is 2.AD, is used. We therefore see that while only two prisms of different materials are necessary to produce a spectrum without deviation of a specified ray, an odd number of prisms is used in practice to obtain greater dispersion and to avoid a lateral shift of the sodium ray.

---

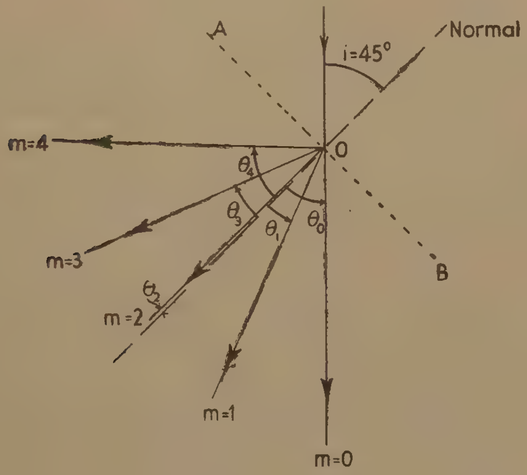
* Amici, Giovanni Battista. This author is quoted frequently as J. B. Amici, but the above name appears in the Royal Society's 'Catalogue of Scientific Papers.' Southall [' Geometrical Optics,' p. 492] states that he was unable to verify the reference given—the same reference appears in Kayser's ' Handbuch der Spektroskopie.' A new search by the author has failed to find a copy of the original paper in this country.

The Plane Diffraction Grating: Principal Maxima.

If m is the order of the spectrum, we have the familiar equation

$$m\lambda=(a+b)\ (\sin i+\sin\theta).$$

Fig. 18.



Frequently the above equation is wrongly given as

$$m\lambda=(a+b)\ (\sin i\pm\sin\theta),$$

for i and θ are chosen positive when they lie on the same side of the normal, and therefore the negative sign must not appear for $\sin\theta$ is an algebraic quantity. If, therefore, $0>\sin\theta>-1$, θ must lie on the other side of the normal than does i (*cf.* the numerical example below).

In experimental work with a plane diffraction grating the angle of incidence is usually made zero: when this is done it is seldom realized that higher order spectra can be obtained by inclining the grating to the incident beam. Thus, if $\lambda=6\times10^{-5}$ cm. and $(a+b)=(6\times10^3)^{-1}$ cm., then the third and higher order spectra are absent if $i=0$. Suppose, however, that $i=45°$. Then the third and fourth order spectra are to be seen, for we have

$$\theta_4=47° 8', \quad \theta_3=21° 54', \quad \theta_2=0° 44', \quad \theta_1=-20° 19', \quad \theta_0=-45°.$$

If we put $m=-1$ and call θ_1' the corresponding angle of diffraction, $\sin\theta_1'>1$: all negative orders are therefore unobtainable. The positions of the principal maxima are shown in fig. 18.

In practice it is desirable to measure the angles of diffraction without reference to the position of θ_0 : with a grating inclined to the incident light at 45° this is possible if the grating is turned through 90° and the various spectra are again observed in symmetrical positions.

Some Common Errors and Pitfalls in Acoustics.

In teaching acoustics a satisfactory proof is usually given that for sound waves of small amplitude travelling in an ideal gas the velocity of propagation, c, is given by

$$c=\sqrt{\frac{\gamma p}{\rho}}=\sqrt{\gamma \mathscr{R}\mathrm{T}},$$

or

$$c \propto \sqrt{\mathrm{T}},$$

where the symbols have their usual meanings. Unfortunately, however, the approximation $c \propto (1+\frac{1}{2}\alpha\theta)$, where θ is the temperature on a centigrade scale, and α the coefficient of increase in pressure at constant volume with temperature, is then introduced. The equation is only valid if $\alpha\theta$ is small ; if it is given, students almost invariably fall into the error of using it on all occasions, *e. g.*, when $\theta=100°$ C. or $\alpha\theta \simeq \frac{1}{3}$. It seems desirable therefore never to introduce the approximation, but to use

$$\frac{c}{c_0}=\sqrt{\frac{\mathrm{T}}{\mathrm{T}_0}}.$$

In laboratories sonometers are often found with wires which are considerably too thick for accurate work. Allan * has shown that steel wire S.W.G. 32 should be used—it must certainly not be thicker. Theory shows that for thick wires a correction involving Young's modulus of elasticity, E, should be made. Since this correction involves E, it seems unscientific to say that the correction is made on account of the rigidity of the material : would it not be preferable to say " because the wire has flexural rigidity " ?

Moreover, the theoretical treatment of the above correction as generally given in text-books is invalid. The reason for this is mainly because ideas on bending moments are not sufficiently definite—attention has already been given to this matter in these notes, and here we have another reason for adopting the sign convention already recommended. Also, it has to be remembered that in dealing with the bending of actual beams the positive direction for the displacement, y, is vertically downwards, whereas in acoustical problems the transverse displacement, ξ, is positive upwards. Hence the bending moment is

---

* Phil. Mag. iv. p. 1324 (1927).

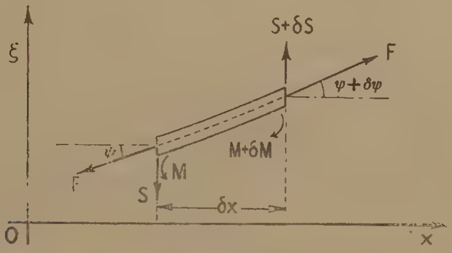
$$\mathrm{EI}\frac{\partial^2 y}{\partial x^2} \quad \text{or} \quad -\mathrm{EI}\frac{\partial^2 \xi}{\partial x^2}.$$

The attempts made by writers to juggle with signs in connexion with this problem would be amusing if their effects were not so tragic!

The derivation of the differential equation for the motion of a stiff wire is as follows:—The system of forces and couples on an element of such a wire, subject to a constant tension F, is shown in fig. 19. As usual, the effects of angular motion will be neglected. The motion at right angles to the x-axis is determined by

$$\mathrm{S}+\delta\mathrm{S}+\mathrm{F}\sin(\psi+\delta\psi)-\mathrm{S}-\mathrm{F}\sin\psi=\mu\,.\,\delta x\,.\,\ddot{\xi},$$

Fig. 19.



where ψ and $(\psi+\delta\psi)$ are the angles indicated, and μ is the mass per unit length of the wire.

$$\therefore \quad \mu\ddot{\xi}=\frac{\partial \mathrm{S}}{\partial x}+\mathrm{F}\cos\psi\,.\,\frac{\partial\psi}{\partial x}$$

$$\doteqdot\frac{\partial \mathrm{S}}{\partial x}+\mathrm{F}\frac{\partial\psi}{\partial x} \qquad [\because\ \psi\to 0 \text{ for small deflexions}]$$

$$=-\mathrm{EI}\frac{\partial^4\xi}{\partial x^4}+\mathrm{F}\frac{\partial^2\xi}{\partial x^2} \qquad \left[\because\ \psi\doteqdot\frac{\partial\xi}{\partial x}\right].$$

$$\therefore \quad \ddot{\xi}=\frac{\mathrm{F}}{\mu}\frac{\partial^2\xi}{\partial x^2}-\frac{\mathrm{E}\kappa^2}{\rho}\frac{\partial^4\xi}{\partial x^4},$$

where $\mathrm{I}=\mathrm{A}_0\kappa^2$, and ρ=the density of the material of the wire$=\mu\mathrm{A}_0^{-1}$. [I is the "moment of inertia" of a cross-section of the wire (area A_0) about its line of intersection with the neutral axis; κ is the corresponding radius of gyration.]

Writing in turn F →0 and E→0, the appropriate equations for the motion of bars and of strings are respectively obtained. The solutions are given in many well-known text-books, so that it is not necessary to repeat them here.

The elementary theory of a Helmholtz resonator shows that its natural frequency, f, is given by

$$f=\frac{c}{2\pi}\sqrt{\frac{\mathrm{S}}{\mathrm{V}l}}=\frac{c}{2\pi}\sqrt{\frac{\kappa}{\mathrm{V}}},$$

where $\kappa=\frac{S}{l}$. S is the cross-sectional area of the neck, l the length of the neck, V the volume of the resonator, and c the velocity of sound in air at the temperature of the experiment. Now if the resistivity (electrical) of the medium (air) were unity, the electrical resistance of the air in the neck would be $1\times\frac{l}{S}$; the electrical conductance would be $\frac{S}{l}$. It is therefore regrettable that κ has been termed the "conductivity" of the neck : a better word is "conductance" and even then the analogy is not too good, for the medium has been given a very special numerical value for its resistivity.

Some Fundamental Definitions in Electricity and Magnetism.

In defining a unit charge in electrostatics, surely it is essential to state the precise nature of the charge : it is therefore proposed that the definition should be as follows :—

The unit charge of electricity is a point charge, of the type of electricity found on glass when this is rubbed with silk and conventionally termed positive, such that, when placed at a distance of 1 *cm. in a vacuum* (*air*) *from an equal point charge, the force of repulsion between the charges is one dyne.*

Then the statement that the charge on an electron is given by

$$e=-4{\cdot}805\times10^{-10}\ \text{e.s.u.}$$

is quite definite.

In dealing with fundamental conceptions and problems in electrostatics (and similarly in magnetism) it is unnecessary and certainly most undesirable to introduce unit charges (or unit poles). It is invariably pointed out that the field is certainly disturbed when such a charge is introduced, but confusion arises because units are suppressed or wrongly used. A better method of approach, even with elementary students, is to introduce a small charge δq. Then if δF is the force which the charge experiences when at a point P in an electric field, then $\frac{\delta F}{\delta q}$ is the electric intensity, E, at P. Strictly speaking we ought to use $\lim\frac{\delta F}{\delta q}=\frac{dF}{dq}=E$, but if δq is sufficiently small the point is not important at early stages. The essential point to emphasize is that when δF is in dynes, δq in e.s.u. of charge, then E is expressed in dynes per unit charge.

A similar method can be used in numerical examples.

Also, if δW is the work done against the field in bringing up a charge δq from infinity to the point P in an electric field, then $\frac{\delta W}{\delta q}$ is the electric potential at P $\left[\text{strictly }\lim\frac{\delta W}{\delta q}\right]$. The unit of electric potential is therefore the dyne per unit charge. If a unit charge is used in this connexion the potential is wrongly defined as the work done against the field in bringing up that charge from infinity to P. This gives the impression that potential may be expressed in ergs. If teachers wish to retain the definition involving a unit charge, then it is only true to state that the potential at P is *numerically* equal to the work done against the field in bringing up the unit charge from infinity to P, it being assumed that the field is not disturbed by the introduction of the unit charge.

Similar definitions are advocated in connexion with magnetic intensity and magnetic potential.

In current electricity the term resistivity is now generally used instead of specific resistance. This is desirable. It seems a pity, however, that the definition of resistivity is not generally such that the units for resistivity are at once apparent. The author uses the following definition:—Consider a uniform wire of length l and cross-section A. If R is the resistance of the conductor, which is nothing more than the drop in potential along the conductor divided by the steady current flowing, then

$$\mathrm{R} \propto \frac{l}{\mathrm{A}}.$$

To get rid of the sign of proportionality we write

$$\mathrm{R} = \sigma \cdot \frac{l}{\mathrm{A}},$$

where σ is termed the resistivity of the material of the conductor. From this we have at once

$$[\text{ohm}] = [\sigma] \cdot \frac{[\text{cm.}]}{[\text{cm.}^2]},$$

so that σ must be expressed in ohm. cm. Again it is seen that σ cannot be measured in ohm. cm.^{-3}, a unit which arises from the confusion caused by defining σ as the resistance of a unit cube. The unit cube and the ideas associated with it ought to be banished from modern teaching: the life of the unit cube has been far too long already!

Then just as the conductance, S, of a conductor is the reciprocal of its resistance, R, the conductivity, κ, is the reciprocal of the resistivity, the unit being the ohm.^{-1} cm.^{-1}—never the reciprocal ohm!

The definition of the Peltier coefficient is usually given badly: it is regrettable that Maxwell's original definition should have been dropped by eminent physicists and these later authorities copied by others. Maxwell * wrote, "*Π being the coefficient of the Peltier effect, that is, the heat absorbed at the junction by unit current in unit time.*" It seems ridiculous to use the expression "*heat absorbed or emitted*" in the definition, as is often done, for the sign of the coefficient is lost, whereas, if we adopt Maxwell's definition, then when thermal energy is emitted Π is essentially negative.

In dealing with the theory of thermoelectricity, the author advocates the use of a slight modification of Maxwell's original definition of the Peltier coefficient, and writes Π_{ab} to denote the heat (ergs) absorbed at the hot junction of a thermocouple when unit quantity of electricity passes from the metal A to the metal B. Passing to the Thomson coefficient, this should be defined as follows:—*Let a charge* δq *be taken from a point at absolute temperature* T *to another at temperature* $T+\delta T$ *in a homogeneous conductor. Then the energy gained by the charge is* $\sigma\delta q\,.\,\delta T$, *where* σ *is the Thomson coefficient for the material of the conductor.* Thus σ is positive when there is an acquisition of energy by the charge under the above circumstances.

---

* 'Electricity and Magnetism,' vol. i. p. 369.

The argument, by means of which an expression for the conductivity (specific conductance) of an electrolyte, first given by Kohlrausch*, is obtained, is seldom set out with anything like rigour. Suppose that a solution contains x gm. of a substance whose molecular weight is M, dissolved in y cm.^{3} of solution. Then the concentration is $\frac{x}{yM}$ mole. cm.$^{-3}$. Call this N mole. cm.$^{-3}$. Suppose that of this number a fraction α is ionized where $0<\alpha<1$, and that each mole which is ionized splits up into two gram-ions each of valency z. Each gram-ion will carry a charge numerically equal to zF, where F is the charge in coulombs carried by a gram-ion of a monovalent substance. Let $|u_+|$ and $|u_-|$ be the mobilities of the cation and anion respectively, *i. e.*, numerically the values of their velocities under a potential gradient of 1 volt cm.$^{-1}$. If the potential gradient is $\left|\frac{\partial V}{\partial x}\right|$ volt. cm.$^{-1}$, then the velocities of the ions will be $|u_+|\,.\left|\frac{\partial V}{\partial x}\right|$ and $|u_-|\,.\left|\frac{\partial V}{\partial x}\right|$ respectively. Hence the quantity of electricity passing per second across a small area, S, normal to the direction of the flow of electricity, is

$$\alpha NzFS\Big[|u_+|+|u_-|\Big]\,.\left|\frac{\partial V}{\partial x}\right|.$$

This is the current I in amperes. Hence

$$\begin{aligned} I&=\alpha NzFS\Big[|u_+|+|u_-|\Big]\,.\left|\frac{\partial V}{\partial x}\right| \\ &=\kappa S\left|\frac{\partial V}{\partial x}\right|, \end{aligned}$$

where κ is the conductivity in ohm^{-1} cm.$^{-1}$.

$$\therefore \quad \kappa=\alpha N\Big[|u_+|+|u_-|\Big]zF \text{ ohm}^{-1}\text{ cm.}^{-1}.$$

The conductivity of a solution divided by the concentration of the solute (in mole. cm.$^{-3}$) is termed the *molecular conductivity* of the solution. The conductivity divided by the number of gram-equivalents per cm.^{3} is called the *equivalent conductivity* (Λ).

$$\therefore \quad \Lambda=\frac{\kappa}{zN}=\alpha\Big[|u_+|+|u_-|\Big]F.$$

---

* Wied. *Ann.* vi. p. 199 (1879).

Thus Λ depends only on the degree of ionization and the ionic mobilities. The unit is the ohm^{-1} cm.$^{-1}$ (gm.-equiv.)$^{-1}$ cm.^{3}, or ohm^{-1} cm.^{2} (gm.-equiv.)$^{-1}$.

The Joule Effect.

If I amps. is the current flowing through a conductor when V volts is the potential difference across it, then the heat developed in it in t seconds is given by

$$W = VIt \text{ joules.}$$

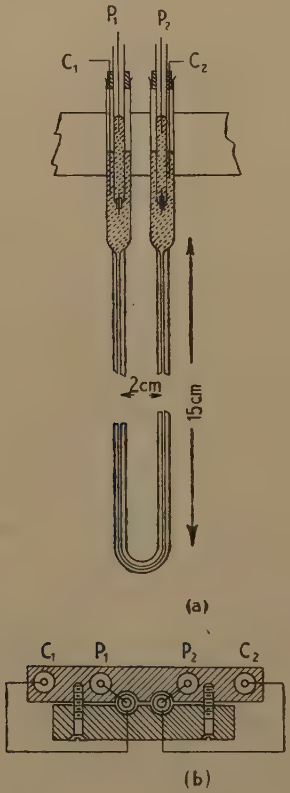
This is often written in one or other of the alternative forms I^2Rt, or $V^2R^{-1}t$, where R ohms is the resistance of the conductor. The use of these alternative expressions is to be deprecated especially in calorimetric work. Callendar always emphasized in his lectures that the value of R under the conditions applicable to a calorimetric experiment was never known unless both V and I were measured. The reason for this is that the temperature of the wire is always in excess of that of the calorimetric fluid by an unknown amount, so that even if the resistance of the wire has been measured at a standard or known temperature and the coefficient of increase of resistance with temperature for the material of the wire is known, it is nevertheless impossible to calculate a value for the resistance of the wire when it is used in a calorimetric experiment. The effect is small if manganin wire is used, but the argument is greatly clarified by the following experiment:—A nickel wire, S.W.G. No. 34, and about 40 cm. long, was used as a heater. Its resistance at room temperature was about 0·85 ohm. It was hoped to use this heater in an experiment to determine the specific heat of paraffin oil, a value for J, the mechanical equivalent of heat, being assumed. When a high resistance voltmeter was placed across the wire, the current through it being about 2 amps., the fraction V/I differed considerably from R, as determined by a P.O. box, even when the liquid was thoroughly stirred: when stirring ceased, the voltmeter reading decreased by as much as ten per cent. Hence, in experiments to determine J electrically or, assuming J, to determine the specific heat of a liquid, it seems most undesirable to use any other expression than V It for the energy dissipated in the calorimeter.

Experimental Determination of the Coefficient of Increase in Resistance with Temperature for Mercury.

The mercury was contained in a pyrex glass capillary U-tube, about 1 mm. in internal diameter, the upper ends being fused to tubes about 5 mm. in diameter. Two corks carried narrow glass tubes through the ends of which tungsten wires were sealed: a small quantity of mercury was placed in each tube and copper wires, arranged as shown in fig. 20 (*a*), served as "potential leads." The "current leads" were also of copper and were wound round the outside of the above narrow glass tubes: in this way sufficient rigidity was secured. [This detail is omitted from the diagram.] Two ebonite blocks, carrying four brass terminals, were placed near the upper portion of the system of tubes and electrical connexions were effected as suggested in fig. 20 (*b*).

To determine the resistance of the mercury column at each of the two steady temperatures used, a Rayleigh potentiometer was employed. A standard 0·1 ohm manganin resistance coil was placed in series with the "mercury resistance," one accumulator and an adjustable resistance. The resistance of the mercury column was measured in the usual way, the total resistance of the Rayleigh potentiometer being 11,000 ohms : the potential difference across it was about 0·3 volt. The current through the mercury in the U-tube was about one ampere, so that it was necessary to investigate whether or not the Joule effect in the mercury caused this to be at a temperature appreciably different from that of the bath in which the apparatus was placed. The current was doubled, *i. e.*, the heating effect was quadrupled, but within the limits of accuracy of the coils used no effect could be detected.

Fig. 20.



It was found that $R_0=0\cdot1418$ ohm, and $R_{100\cdot0}=0\cdot1558$ ohm, so that the temperature coefficient required is

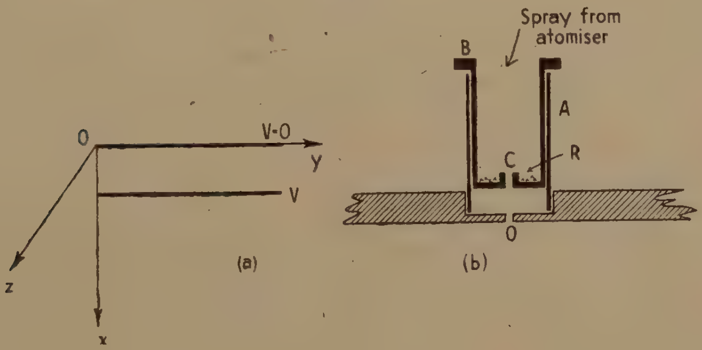
$$\alpha=9\cdot95\times10^{-4}\ \text{deg.}^{-1}\ \text{C.},$$

a value which only differs by 0·5 per cent. from that given in the International Critical Tables *.

Theory of Millikan's Experiment to Determine the Electronic Charge.

By treating all quantities concerned algebraically much confusion in connexion with the theory of this well-known experiment may be avoided. Three rectangular axes are selected and arranged as shown in fig. 21 (*a*), so that Oy coincides with the trace of the upper plate of the condenser on the plane xOy. [The usual right-handed system of axes has been chosen.] We shall consider all points in the plane yOz, *i. e.*, the plane containing the upper plate of the condenser, to be at zero potential and let V e.s.u. denote the potential elsewhere. [In the actual experiment the upper plate is not permanently earthed.] The intensity of the electric field in the region below the aperture in the upper plate [not shown] is $-\frac{\partial V}{\partial x}$. If q is the charge in e.s.u. (strictly c.g.s.e.s.u.) on the drop examined—[note that q is an algebraic quantity]—the force on the charged drop due to its presence in the electric field is

Fig. 21.



$$q\left[-\frac{\partial V}{\partial x}\right],$$

and this force is considered positive when it acts in the direction of x increasing, *i. e.* downwards. Let us assume that the motion of the drop is steady, *i. e.*, the drop moves without acceleration, so that the resultant force on the drop due to gravity and the electric field must be equal and opposite to the force arising from its steady motion in a viscous medium. If m is the effective mass of the drop, g the intensity of gravity, and if Stokes' law is valid for such small drops, we have

$$mg-q\frac{\partial V}{\partial x}=6\pi r\eta(\dot{x})_f, \quad \ldots \ldots \quad (1)$$

---

* Vol. vi. p. 136.

where r is the radius of the drop, η the coefficient of viscosity of air, and $(\dot{x})_f$ is the velocity of the drop in the positive direction of x when the electric field is applied and the motion is steady. When the electric field is off, we have

$$mg = 6\pi r\eta(\dot{x})_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where $(\dot{x})_0$ is the steady velocity of the drop when the electric field is zero. From (1) and (2) we have

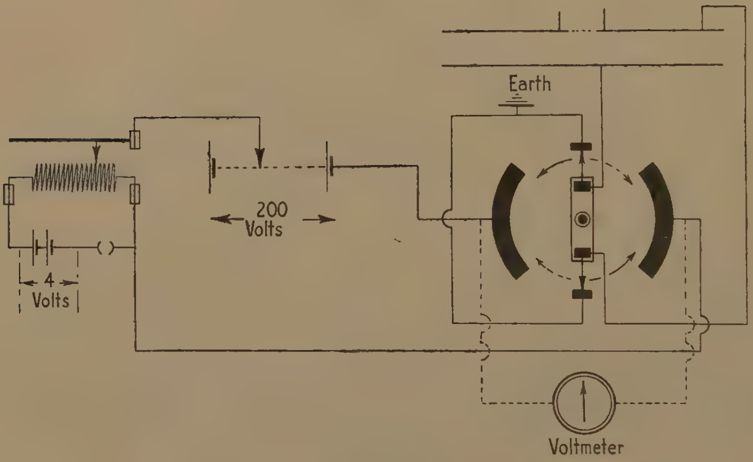
$$q = \frac{mg\left[1 - \frac{(\dot{x})_f}{(\dot{x})_0}\right]}{\frac{\partial V}{\partial x}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

so that q can be calculated when the mass of the drop is known.

From (2) we have

$$(\dot{x})_0 = \frac{2}{9}\,\frac{gr^2(\rho - \rho_0)}{\eta},$$

Fig. 22.



which gives r providing that ρ and ρ_0, the densities of the oil and of the air under the conditions of the experiment, are known.

The value for the electronic charge is deduced in the usual manner.

Millikan's Experiment as a Laboratory Exercise.

When this experiment is performed it is inconvenient to have a potential difference across the condenser much in excess of 200 volts. This means that the charge on the drop must consist of several electrons if the motion of the drop is to be vertically upwards when the electric field is applied. Difficulty, probably increased in damp weather, is sometimes experienced in getting a sufficiently large charge on the droplet. This can be avoided in the following way:—The upper plate of the condenser consists of a piece of brass about 5 mm. thick. At the centre of this plate a hole, about 1 mm. in diameter, is drilled; a deep recess is then made in the brass so that the above aperture is now at the centre of its base. Into this recess a brass tube, A, fig. 21 (*b*), about 2 cm. long, is soldered. B is a brass tube sliding in A and it has a collar to support it in position. A smaller tube, C, is soldered into the base of B, and a small quantity of radioactive material (cleveite is quite good enough for this purpose) is placed in the annulus between B and C so that the air through which the droplets fall before entering the experimental chamber is ionized. Of course X-rays may be used for the same purpose—as, in fact, was done by Millikan—but their use requires an unnecessarily elaborate arrangement for ordinary laboratory work.

A convenient set-up of the apparatus for a determination of the electronic charge is shown in fig. 22.

Acknowledgment.

In conclusion, the author would like to express gratitude to Prof. Allan Ferguson, D.Sc., who suggested that this paper should be written, for his helpful discussion on several of the topics raised; also to Prof. F. Horton, F.R.S., who has readily granted permission for the facilities of this laboratory to be used in connexion with the construction of new apparatus.

LXXXIV. *Probability and Quantum Theory.*

By Harold Jeffreys[*].

[Received August 10, 1942.]

1. Modern quantum theory is stated in many forms, which are generally recognized to be equivalent; that is, when observable predictions are made they make the same prediction. In other words, the different forms of the theory must be regarded as representing different choices of language; they say the same things about observations in different ways. But the choice of a language for scientific purposes is not wholly arbitrary. Its rules must be framed so that in some way they express the fundamental laws of the system, which are relations between observations. If they express more than that, the additional statements are either verifiable, in which case they await verification, or non-verifiable and superfluous. The various theories are often stated to deal wholly with observable quantities, and the results are usually stated in terms of probabilities. But observable quantities are mathematically real, and probabilities real and

---

[*] Communicated by the Author.

non-negative. It is therefore somewhat puzzling to find that the fundamental statements of the theories involve complex numbers, non-commutative multiplication, or matrix roots of -1. Real combinations of these are identified with observables, and the probabilities calculated are positive, so that the conditions required are satisfied, but the logical status of the complex quantities remains mysterious. If they are really fundamental the world needs many additional dimensions to describe it. If they are merely introduced for mathematical convenience, it should be possible to show what there is in the laws connecting observations that makes their introduction convenient. An illustration is provided by the complex potential $V+iU$ in two-dimensional electrostatics. This is not usually regarded as having any direct physical meaning. Its usefulness rests on the fact that electric force is derivable from a potential satisfying Laplace's equation, and any real function satisfying this equation can be proved to be the real part of a function of the complex variable $x+iy$. Thus the introduction of the complex variable is in this case demonstrably convenient. But without such a demonstration, resting ultimately on the inverse square law, two-dimensional potential theory would be a language without any clear reason for using it. It is not clear what there is in the physical laws that makes the complicated languages of quantum theory convenient to express it.

I shall assume that the fundamental problem is that of prediction, that is, given information from observations of a system at one instant, to infer what we can about its state at another instant. In classical mechanics the position of a system is specified by a set of co-ordinates, but we know that the co-ordinates by themselves at one instant lead to no useful prediction about their values at another; we need also their initial rates of change or the corresponding momenta. So much must also be true in quantum mechanics. For a system with n degrees of freedom we need information about $2n$ variables at an instant before we can make a prediction. These variables must be real quantities. The information about the state of the system at any instant will then be expressed as a joint probability distribution of $2n$ real variables. We shall require for completeness that no relevant quantities for the system are omitted from the list of $2n$ that are specified. These variables are not exactly known at any one instant, on account of observational error. If different co-ordinates are observed at different instants there will be additional complications, but it remains true that the statement of the problem requires a probability distribution for $2n$ variables. We may, however, restrict ourselves at present to the case where all the data refer to observations at the same instant. If we know the intrinsic properties of the system (in C.M. the equations of motion), it will then be possible to calculate the joint probability, given the initial conditions, of all variables at any other instant; and to get the probability distribution for any one variable we shall have only to add or integrate with regard to the others. The possibility of prediction at all involves as a fundamental idea that of this joint distribution at time t; and we may take the data H to be whatever information we may have about the state at time 0, together with any admissible form of the equations of motion or whatever will replace them in the quantum scheme. Without attempting at present to specify the meanings of the variables precisely, we may call them provisionally the co-ordinates q and momenta p. Then, if $\mathrm{P}(dq'dp' \mid \mathrm{H})$ is the probability that the various q's and p's will lie in particular ranges $dq'dp'$ at time t, $\mathrm{P}(dqdp \mid \mathrm{H})$ the probability, supposed derived from observation, that they lie in particular ranges $dqdp$ at time 0, $\mathrm{P}(dq'dp' \mid qp\mathrm{H})$ the probability that they will lie in the ranges $dq'dp'$, given both H and the statement that they are in the ranges $dqdp$ at time 0, we have by the product rule for probabilities of joint assertions (Jeffreys, 1939, p. 24),

$$\mathrm{P}(dq'dp'dqdp \mid \mathrm{H}) = \mathrm{P}(dq'dp' \mid qp\mathrm{H})\mathrm{P}(dqdp \mid \mathrm{H}), \quad . \quad . \quad . \quad (1)$$

and integrating over q and p (that is, applying the addition rule for probabilities to the disjunction of the mutually exclusive propositions that q and p lie in the various different ranges possible)

$$\mathrm{P}(dq'dp' \mid \mathrm{H}) = \int_q \int_p \mathrm{P}(dq'dp' \mid qp\mathrm{H})\mathrm{P}(dqdp \mid \mathrm{H}). \quad . \quad . \quad . \quad (2)$$

Some features of this formula should be noticed at once in relation to general probability theory. It is understood that dq and dp under the sign P( ) are abbreviations for propositions, e.g., dq means the proposition that q lies in a range from say q_1 to q_1+dq, where q_1 can be replaced by q in further writing. In the data qpH, q and p are really also abbreviations for the propositions that q and p lie in particular ranges dq, dp; but it is convenient to abbreviate these propositions in different ways according as they appear in the data or in the proposition whose probability with respect to those data is being considered. The reason is that in such formulæ as (2) $\mathrm{P}(dqdp \mid \mathrm{H})$ is an element of a distribution, usually continuous, and the differentials in its expression call attention to this fact; but in $\mathrm{P}(dq'dp' \mid qp\mathrm{H})$ the variation of q and p in elementary ranges ultimately contributes nothing to the sum, and we need attend only to the values of q and p.

In classical mechanics $\mathrm{P}(dq'dp' \mid qp\mathrm{H})$ is a point concentration; that is, it is 1 if $dq'dp'$ includes the values at time t that would be inferred from the rules of classical mechanics given the values q and p at time 0, and otherwise is 0. If point-concentrations occur the integrals may be replaced by Stieltjes integrals or the δ-function may be introduced.

It may appear at first sight that point-concentrations will not occur in quantum theory, but this is not obvious. There is, so far as I know, only one case where a solution by the Schrödinger method has given a probability distribution the data for which can be easily specified. The Darwin solution (1927) for the free electron in a uniform field may be written

$$\mathrm{P}(dx' \mid \mathrm{H}) \propto \exp\left[-\frac{(x'-x_0-ut-\frac{1}{2}gt^2)^2}{2\{\sigma^2+(ht/4\pi m\sigma)^2\}}\right]dx'. \quad . \quad . \quad . \quad (3)$$

But if we take $$\mathrm{P}(dx \mid \mathrm{H}) \propto \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}\right\}dx, \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$P(dp\,|\,H)\propto \exp\left\{-\frac{(p-mu)^2}{2(h/4\pi\sigma)^2}\right\}dp, \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

and regard the uncertainties of x and p as independent, we have

$$P(dxdp\,|\,H)\propto \exp\left\{-\frac{(x-x_0)^2}{2\sigma^2}-\frac{(p-mu)^2}{2(h/4\pi\sigma)^2}\right\}dxdp. \quad . \quad . \quad (6)$$

If we also take

$$P(dx'dp'\,|\,x, p, H)=\delta\left(x'-x-\frac{pt}{m}-\tfrac{1}{2}gt^2\right)\delta(p'-p)dx'dp', \quad . \quad . \quad (7)$$

form $P(dx', dp'\,|\,H)$ by (2), and integrate with regard to p' to obtain the probability distribution of x' by itself, we get (3). Thus the whole uncertainty of the prediction is traceable to that of the initial conditions; the dynamics of the system is contained in K, and there is nothing in the solution to indicate that this is less deterministic than in classical mechanics.

It seems best to leave open at present the question whether K can be a point-concentration in quantum mechanics. But the statement that a joint probability-distribution of q and p exists conflicts with some statements of the uncertainty principle, which would deny that the proposition that q and p lie in particular ranges $dqdp$, which may be indefinitely short, has any meaning at all. These statements, however, seem to me to miss the point of the principle and to have misleading analogues in other fields. It is loosely said, that is, that q and p, if conjugate, are not simultaneously measurable. But the whole of the arguments for the uncertainty principle go to show that they can be simultaneously measured. These arguments do show that there is a lower limit to the product of the uncertainties attainable, but this is fully expressed by representing the information with respect to their values by a joint probability distribution around a pair of estimates, which are real quantities. The fact that we could not in practice know that q and p lay in indefinitely narrow ranges $dqdp$ does not in the least prevent us from considering the proposition that they do; to say that it does assumes that we cannot attach a meaning to a proposition without pointing to a case where it could be known with certainty. But that would invalidate the whole theory of measurement; we can never in practice say that a measured quantity is certain to lie within any particular range unless that range is between $\pm\infty$. Similarly, we could not consider the consequences of an atom being in a steady state without being able to say certainly that some particular atom is in a steady state. The argument confuses asserted and considered propositions, and it is precisely in order to deal with considered propositions that the theory of probability exists. The matter has been discussed also by A. J. Ayer (1936, p. 214), speaking on Mach and his followers: " They require of a synthetic proposition that it should, in principle at least, be conclusively verifiable. And as, for reasons which we have already given, no proposition is capable, even in principle, of being verified conclusively, but only at best of being rendered highly probable, the positivist criterion, so far from marking the distinction between literal sense and nonsense, makes every utterance nonsensical. . . . It is necessary . . . . to allow a proposition to be genuinely factual if . . . . it is capable of being substantiated to any degree whatsoever by any possible observation. In practice, indeed, very little of what is allowed to be significant by this criterion would not be accepted by the positivists. But that is because they do not apply their own criterion consistently." On the other hand, it is certainly meaningless to speak of a probability without specifying the data, and though functions obtained in quantum theory are interpreted as probabilities, there are hardly any cases where the data are properly specified.

The p's are certain real quantities needed to describe the motion ; they are not identical with any q-numbers or matrices*, and need not be identical with the classical momenta or have any intimate relation to mdx/dt. The latter may not exist, and an attempt to observe it directly would evidently give $\pm\infty$ if it could be carried out, on account of the finite uncertainty of x at each end of the short interval dt.

We assume throughout that the system is left to itself from time 0 up to the instant when the probability distribution is required. It is therefore essential that the system shall not be re-observed at any intermediate instant. If it is to be re-observed at the end, the prediction of the observations should take account of the deviations that will be introduced at the second observation. But these have nothing to do intrinsically with the length of the interval, and we are entitled to consider the unobserved system as a standard on which they will be superposed. It would not be valid to apply the same treatment to a system continually observed, for the intermediate disturbances would lead to a very different result from that for the system left to itself.

Then we can write

$$\mathrm{P}(dqdp \mid \mathrm{H}) = \mathrm{F}(q, p)dqdp \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

$$\mathrm{P}(dq'dp' \mid \mathrm{H}) = \mathrm{G}(q', p')dq'dp' \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

$$\mathrm{P}(dq'dp' \mid qp\mathrm{H}) = \mathrm{K}(q', p', q, p, t)dq'dp', \quad . \quad . \quad . \quad . \quad (10)$$

and call F, G, K probability densities. All are essentially non-negative. Then (2) becomes

$$\mathrm{G}(q', p') = \iint \mathrm{K}(q', p', q, p, t)\,\mathrm{F}(q, p)dqdp. \quad . \quad . \quad . \quad . \quad (11)$$

This equation gives us a formal rule for finding the probability distribution for the co-ordinates and momenta at time t, given that at time 0, and also the properties of the system. Its nature is epistemological, apart from the fundamental physical assumption that nothing is relevant to $(q'p')$ except (q, p), and it can give no actual information about observable quantities until we have information to fix F and K. On the other hand, any suggested F, G and K must be consistent with it. Alternatively, given G, it could be regarded as an integral equation to find F, K being the kernel. The whole of the mechanics of the system is contained in K. K is, in general, unsymmetrical in the accented and unaccented letters. For in classical mechanics it is a point-concentration at the values of q' and p' that q and p would lead to in time t; but $K(q, p, q', p', t)$ would be a point-concentration at the values of q' and p' that would lead to q and p in time t, and these values are not the same.

---

* It is really rather unfortunate linguistically that Heisenberg and Dirac call their q-numbers by the preoccupied names of co-ordinates and momenta. It would be better to call them skewordinates and skewmenta and reserve the older names for the real quantities determined, with specifiable uncertainties, by observation.

We shall make use of two elementary physical principles, suggested by classical mechanics but capable of being stated in probability language in much more general conditions, namely the relativity of time and the reversibility of path. According to the former, for any initial instant t_0 the same K is needed to give the probability distribution at time t_0+t, if t is the same. Consequently if we take

$$H(q'', p'')=\iint K(q'', p'', q', p'\ t)\ G(q'\ p')dq'dp', \quad . \quad . \quad . \quad (12)$$

$H(q'', p'')$ will be the probability density for the variables at time $2t$, and so on indefinitely. The probabilities after successive intervals t may then be regarded as connected by a chain of linear transformations. The theory of transformations in chains of this type has been investigated by several different methods by Poincaré, Frobenius, Markoff, Kolmogoroff, Fréchet and others; a full account is given by Fréchet (1938)*.

If we take a classical system at time t and reverse all the momenta, the system will retrace its path. In time t it will therefore reach the state originally described at time 0, except that all the momenta will be reversed. In our generalized scheme we have not yet defined the momenta. A sequence of probability distributions for the co-ordinates alone, varying with time, can be stated. We assume that n parameters p exist such that if they are reversed at time t the probability distributions for the q's will pass through the same sequence in the reverse order. This being supposed possible at any time, we have

$$F(q', -p')=\iint K(q', p', q, p, t)\ G(q, -p)dqdp, \quad . \quad . \quad . \quad (13)$$

or, by a change of variables,

$$F(q, p)=\iint K(q, -p, q', -p', t)\ G(q', p')dq'dp'. \quad . \quad . \quad (14)$$

This may be regarded as the solution of (11).

Alternatively, if we take the distribution at time t and use the same q', p' but interval $-t$ instead of t, we should get back to $F(q, p)$; hence

$$F(q, p)=\iint K(q, p, q', p', -t)\ G(q', p')dq'dp', \quad . \quad . \quad . \quad (15)$$

and by comparison with (14), if this is to hold for all G,

$$K(q, p, q', p', -t)=K(q, -p, q', -p', t), \quad . \quad . \quad . \quad (16)$$

This may be regarded as a definition of K for negative t.

---

* I am much indebted to Mr J. H. Moyal, who gave me this reference.

From the rule that 1 denotes certainty on the data, we have

$$\iint F(q,p)dq\,dp=1\,;\quad \iint G(q,p)dqdp=1\,;\quad \iint K(q',p',q,p)dq'dp'=1\,; \quad (17)$$

and applying the same rule to (14)

$$\iint K(q',p',q,p)dqdp=1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

We shall find reason later to make a small but important change in this hypothesis ; meanwhile, we shall follow out its consequences.

It may be thought that reversing the momenta would entail an additional Heisenberg uncertainty, and that therefore in quantum theory the final F(q, p) would not be the same as the initial one. But this supposes the reversal experimental. The reversal here is not an experimental one ; we are considering an idealized case, not the complications introduced by the practical attempt to achieve it. The statement holds, for instance, for the Darwin solution for the free electron.

Now if we can find a set of functions $F_l(q, p)$ such that any physically possible F(q, p) can be expressed linearly in terms of them, and such that

$$\iint K(q', p', q, p, t)\, F_l(q, p)dqdp=\lambda_l F_l(q', p'), \quad . \quad . \quad . \quad (19)$$

where the λ_l are functions of t only, we can take

$$F(q, p)=\Sigma A_l F_l(q, p)\,; \quad G(q', p')=\Sigma \lambda_l A_l F_l(q', p'), \quad . \quad . \quad (20)$$

$$H(q'', p'')=\Sigma \lambda_l^2 A_l F_l(q'', p''), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

and so on, by the relativity of time. This does not assume reversibility. The solution will then be formally complete if we can solve this eigenfunction problem.

All information relevant to observation is contained in G(q', p') ; if two forms of K lead to the same λ_l and F_l the difference between them is therefore conventional. Now if the F_l are normalized, have complex orthogonality, and can be found otherwise, a suitable form of K will be

$$K=\Sigma \lambda_l F_l(q', p')\overline{F}_l(q, p),$$

and

$$A_l=\iint F(q, p)\overline{F}_l(q, p)dqdp.$$

It may happen that with some forms of K there will be eigenfunctions that will affect no observation ; if so, it is irrelevant to physics whether these functions are included in the series or not. Such functions may, for instance, be those whose inclusion would represent an uncertainty of q', p' smaller than any observational determination could give, and might be omitted. But in the present approach it is desirable to make no such restriction ; with too few eigenfunctions we might find ourselves unable to *say* physically important things in the language available, but no harm can come from having too many.

2. Fréchet's discussion deals only with the finite case, but we can begin by reducing the general problem to this case. For it will always be possible to sub-divide the co-ordinate-momentum space into a finite number of sub-ranges, the extents of which are arbitrary except that they must

together cover the whole range of possible values. If there are n particles and we break up the whole range for each q and each p into N, there will be N^{6n} mutually exclusive sub-ranges. Denote this number by M. Even if the range of possible values is infinite this will be possible, since there is no obligation at present to take the sub-ranges all equal. Then if x_j is the probability that the initial values lie within a particular set of sub-ranges indicated by the suffix j, and y_i the probability that those after time t lie within the set indicated by the suffix i, a_{ij} the probability that they will be in the i state at time t given that they are in the j range at time 0, we have the set of M linear equations

$$y_i=\Sigma_j a_{ij}x_j. \qquad (1)$$

Since some state i must occur,

$$\Sigma_i a_{ij}=1. \qquad (2)$$

We have not necessarily

$$\Sigma_i a_{ij}=1, \qquad (3)$$

since this is the sum of probabilities of the same proposition on different data. But our choice of sub-ranges is arbitrary, and we can adjust them so that these relations also hold. According to 1(18) they will hold, but at present we are not assuming reversibility. (They seem to be assumed without statement in several of Fréchet's proofs, though the condition appears previously on his p. 38.) Then we are led to the study of the solutions λ_l, $F_{l,i}$ of the equations,

$$\lambda x_i=\Sigma_j a_{ij}x_j. \qquad (4)$$

If then

$$x_i=\Sigma_l A_l F_{l,i},\ y_i=\Sigma_l\Sigma_j A_l a_{ij}F_{l,j}=\Sigma_l\lambda_l A_l F_{l,i}. \qquad (5)$$

Fréchet's work deals entirely with finite sets of equations and his results are directly adaptable to this set. There is, however, a difference in the point of view, since in his problem the a_{ij} are taken as given; here they depend ultimately on the quantum form of the laws of motion. The latter is not known, and our chief question is to say what properties the a_{ij} must have in order that the solutions may have the properties indicated by current quantum theory.

The equation for λ is $\qquad \| a_{ij}-\lambda\delta_{ij} \|=0. \qquad (6)$

Adding all the rows we have a row all of whose elements are $1-\lambda$; hence 1 is always an eigenvalue. The corresponding $F_{l,i}$ is $1/M$.

For a given λ_l, let the $F_{l,i}$ of largest modulus be such that $|F_{l,i}|=R$. Then for each i (Fréchet, p. 106)

$$(\lambda_l-a_{ii})F_{l,i}=\sum_{j\neq i} a_{ij}F_{l,j}, \qquad (7)$$

$$|\lambda_l-a_{ii}|\,|F_{l,i}|\leqslant R\sum_{j\neq i} a_{ij}=R(1-a_{ii}). \qquad (8)$$

But for one $F_{l,i}$, $|F_{l,i}|=R$; hence for this i,

$$|\lambda_l-a_{ii}|\leqslant 1-a_{ii}. \qquad (9)$$

If then $a_{ii}>0$, λ_l lies within or on a circle with centre a_{ii} touching the unit circle at $+1$. Further, such a circle centred on the smallest a_{ii} will include all the λ_l. Even if all the a_{ii} are 0, no λ_l can have modulus >1.

Evidently, if all the λ_l have modulus <1 except $\lambda=1$, the probabilities will tend to constant limits as the number of applications of the transformation is increased (*i. e.*, as the interval of time considered becomes longer). If there is only one root equal to 1, the limit will be $1/M$ for all i. By an argument due to Markoff (Fréchet, pp. 26–32), if a number ν exists such that after ν stages there is a positive probability that the system will be in any state i, whatever state it was in initially, then there will be a limiting probability distribution independent of the initial one. This result does not depend on (3) ; but if (3) also holds the limit is unique and equal to $1/M$ for all i. This condition is necessary and sufficient (Fréchet, p. 39). By analogy with Liouville's theorem, it is natural to identify it with the condition that the sub-ranges shall be of equal volume in the phase space, and the limiting state with one of uniform probability density, that is, with complete lack of information about the state of the system, subject possibly to the values of functions of the q, p that must remain constant throughout the motion.

The result can be obtained in another way. Suppose that one or more of the λ_l have modulus >1 ; let the largest modulus be μ (possibly repeated). The corresponding terms in x_i cannot cancel for all i, except for special choices of the A_l ; hence by repeating the transformation sufficiently often the probability for some range i will become >1, which is impossible. Hence no $|\lambda|>1$.

Now consider the consequences of the further postulate of reversibility. If we pass backwards in time the solution is obtained by successively dividing the coefficients of the solutions by λ_l. But the states passed through are the same, except for reversal of the momenta, as would be reached by proceeding forwards in time with the initial momenta reversed. But the latter problem has the same eigenvalues and solutions λ_l, $F_{l,i}$; hence, if λ_l is an eigenvalue, $1/\lambda_l$ is another, and the corresponding eigenfunctions differ only by the reversal of the momenta. But if any $|\lambda_l|<1$, this would imply $|1/\lambda_l|>1$, and $1/\lambda_l$ could not be an eigenvalue. Hence the relativity of time and reversibility of path, in the senses defined above, are sufficient conditions for all the $|\lambda_l|$ to be 1, and a_{ij} is a unitary matrix.

At this point we have a possible comparison with the Schrödinger solutions. If we add the probabilities at any time for all sub-ranges of the p's we get the probability distribution at that time for the q's alone, and this will consist of a number of terms each with a time factor of the form $e^{i\sigma t}$. But the Schrödinger solution $|\psi|^2$ for the probabilities of the co-ordinates is of this form when not independent of t, and so far the result is satisfactory.

Now consider the eigenfunction for a given λ_l of modulus 1 but $\neq 1$. Let the largest $|F_{l,i}|=R$. We have

$$\lambda_l F_{l,i} = \Sigma_j a_{ij} F_{l,j} = c, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

say ; whence, taking moduli,

$$R = |\Sigma_j a_{ij} F_{l,j}| = |c|. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

But since $\Sigma_j a_{ij}=1$, $a_{ij}\mathrm{F}_{l,j}$ is a weighted mean of the $\mathrm{F}_{l,j}$, and all their moduli are $\leqslant\mathrm{R}$. Thus all the $\mathrm{F}_{l,j}$ for different j lie within or on the circle $|z|=\mathrm{R}$. A weighted mean of them will therefore lie within the circle and not satisfy $|c|=\mathrm{R}$ unless all $\mathrm{F}_{l,j}$ with non-zero a_{ij} are equal to c. Also since $\lambda_l\neq 1$, $\mathrm{F}_{l,i}\neq c$ by (10), and therefore a_{ii} (not summed)$=0$. Hence there is at least one j other than i such that $|\mathrm{F}_{l,j}|=\mathrm{R}$, and $\mathrm{F}_{l,j}=c=\lambda_l\mathrm{F}_{l,i}$. Also it is impossible, since $a_{ii}=0$, for the system to remain in the sub-range i if it was in that sub-range initially.

Now since $|\mathrm{F}_{l,j}|=\mathrm{R}$ we can repeat the argument with j instead of i; and there will be a k such that $\mathrm{F}_{l,k}=\lambda_l\mathrm{F}_{l,j}$, and $a_{jk}=0$ except for $\mathrm{F}_{l,k}$ with this value. If $k=i$, then $\lambda_l=-1$. If not, we can find another suffix such that we can repeat the process. We cannot proceed indefinitely since the whole number of suffixes is M; hence the sequence must close after a number of repetitions $\leqslant\mathrm{M}$, say m, and then $\lambda_l^m=1$.

If $m<\mathrm{M}$, there will be other values of λ. If these also all have modulus 1, there will be other sets of eigenvalues of the same form; let the numbers in the sets be m, m'. . . Then $mm'm''$ . . . is finite, equal to A, say, and all the solutions are strictly periodic in time At. For each such set there is one value of λ equal to 1.

Now suppose if possible that two values $\mathrm{F}_{l,j}$, $\mathrm{F}_{l,j'}$ are both equal to c. Then a_{ij} and $a_{ij'}$ may both be >0. But if the system was in sub-range i at time 0, it is certain that whatever state the system is in at time $(\mathrm{A}-1)t$ it will be in sub-range i at time At. Hence $a_{is}=1$ for all non-zero x_s at time $(\mathrm{A}-1)t$. But if two states were possible at time $(\mathrm{A}-1)t$ this would make $\Sigma_s a_{is}=2$, and we have a contradiction. Hence only one sub-range is possible at time $(\mathrm{A}-1)t$ given that it was in sub-range i at time 0. Proceeding backwards, we infer that for each sub-range i of the sequence there is one a_{ij} equal to 1 and the rest are 0. We thus have the theorem: *if all the* λ_l *have modulus* 1, *then if the system starts in a particular sub-range it will necessarily proceed in time* t *to another particular sub-range, and will return exactly to the initial state after a finite number of repetitions.* The λ_l must form sets of the form $\lambda_l=\exp 2\pi il/m$; if $m<\mathrm{M}$ the sub-ranges fall into sets such that if the system is in one set it will never pass out of that set, but if $m=\mathrm{M}$ it will pass through the whole set of sub-ranges in time Mt.

With a suitable arrangement of the suffixes, the matrix for $m=\mathrm{M}=2$ is $\begin{pmatrix}0&1\\1&0\end{pmatrix}$, and the determinantal equation $\lambda^2-1=0$. For $m=\mathrm{M}=3$ it is $\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}$, and the determinantal equation $\lambda^3-1=0$.

This result does not require the full hypothesis of reversibility, but only the consequence of it that is already expressed in the time factors of the Schrödinger type of solution. Accordingly, it appears that if these time factors are applicable to the probabilities of finite sub-ranges, the quantum theory is deterministic in the classical sense; any indeterminacy arises from imperfect knowledge of the initial state, and not from the laws of motion.

The highest period corresponds to $l=1$; then $l=2m-1$ is another value, and if $l=1$ gives $\lambda_l=e^{i\sigma t}$, $l=2m-1$ will give $\lambda_l=e^{-i\sigma}$. Now if we again consider the probability distribution for the q's alone, they also are concentrated in one sub-range; and the repetition of the argument with this value of λ_l shows that the system can traverse the same sequence in the reverse order. Hence the hypothesis that all the $|\lambda_l|$ have modulus 1 is a necessary and sufficient condition for reversibility.

If we are gven that a system is in a particular sub-range, the ratios of the coefficients A_l are determinate. If we merely know that it is in some sub range of a cycle, the probability is uniform over all sub-ranges in that cycle. In this case the A_l are all zero except that corresponding to $\lambda=1$. Other ratios of the A_l will correspond to different degrees of ignorance about where the system was in the cycle at time 0.

If a system consists of two independent sub-systems, capable of cycles of m and n stages respectively, the product of every pair of eigenfunctions of the sub-system is an eigenfunction of the joint system; if m and n are prime to each other the joint period is mnt. But here $m+n$ eigenfunctions are sufficient to determine mn. Any two probability distributions for the two systems can be combined by multiplication, so that only $m+n$ coefficients can be assigned independently.

3. Now reversibility in the sense of this paper fails even in classical mechanics if finite fixed sub-ranges of the q's and p's are used. For if we are told only that a system is within a certain range, then since the rate of change of q is always correlated with p, and that of p is usually correlated with q, different values of q and p within a sub-range will lead to positions in different sub-ranges at time t. It is to this fact that the conditions contemplated in the kinetic theory of gases are due (Jeffreys, 1937). The equations of motion contemplate the state as precisely specifiable, but the observational information gives only, say, the mean thrust on a finite area. There is a tendency to reduce by averaging any inequalities in the probability density, and this corresponds to the existence of values of λ with moduli <1. But if we start from the same state with the momenta reversed we again get the same sort of mixing; there will again be values of λ with moduli <1, but in neither case can there be any with moduli >1. Classical reversibility therefore cannot hold unless the state of the system is specified exactly, and not merely by the statement that the co-ordinates and momenta lie within finite ranges, however small. In quantum theory we must therefore expect reversibility to fail if the sub-ranges are taken too large; but we might expect it to hold if the product of the ranges of each pair of conjugate co-ordinates and momenta is less than h. In any case if the sub-ranges are too large the Schrödinger form will fail on account of some of the time factors having modulus <1. We therefore are limited to considering the possibility that this condition will be satisfied if the sub-ranges are small enough.

Now the time interval t is arbitrary, and the only restriction so far made on the extent of the sub-ranges is that $\Sigma_j a_{ji}=1$ for all i, this restriction having the possible interpretation that all the sub-ranges have the same volume in the (q, p) space. The period mt, however, must be an absolute constant of the system. But m is an integer limited by the number of sub-ranges. Hence it is only for a discrete set of values of t that the state of the system can be predicted at all if we keep the number of sub-ranges finite ; we have a quantization of time. The interpretation of λ as $\exp(i\sigma t)$ with continuous time therefore presupposes that the sub-ranges are infinite in number and indefinitely short if the system is confined in a finite volume of the phase-space ; if the system is unlimited they must be infinite in number in any case if they are all of equal volume. This also follows from $a_{ii}=0$; for if we kept the sub-ranges of q and p constant and reduced t, then for sufficiently small t the q's and p's could stay in the initial sub-ranges. The continuous case may be regarded as the limit of a series of discrete ones each satisfying the conditions used. There is then not much hope that the wave mechanics will be valid unless we can speak of the initial co-ordinates and momenta as exactly specified.

The argument may be adapted now to the treatment of the original form 1(19). If the largest value of $|\mathrm{F}_l|(q, p)$ for varying q, p is R for $q=q_0, p=p_0$, we can infer that there is at least one other pair (q_1, p_1) such that $|\mathrm{F}_l(q_1, p_1)|=\mathrm{R}$, and that $\lambda_l \mathrm{F}_l(q_0, p_0)=\mathrm{F}_l(q_1, p_1)$. For all q', p' not satisfying this condition $\mathrm{K}(q', p', q_0, p_0, t)=0$. Thus either $\mathrm{F}_l(q, p)$ is constant over finite ranges specifiable independently of the particular eigenfunction or K is zero almost everywhere. Now all verifiable information about the state of the system at time t is summed up in $\mathrm{G}(q', p')$, and therefore ultimately in the F_l. If different forms for K lead to the same λ_l and F_l the differences between them are conventional. In the former case we can satisfy the conditions by choosing a one-one correspondence between pairs of values (q, p) $(q'p')$ within each pair of ranges where $\mathrm{K}\neq 0$, and taking a finite constant concentration of K at pairs belonging to this correspondence. Thus, in any case we may take K for given $(q\,p)$ to be 0 for almost all $(q'p')$, but to have a finite concentration at corresponding values. Hence we can still speak of determinate paths such that for a given $(q\,p)$ at time 0 the system can describe one or other of them exactly. Since the number of possible states is no longer finite it is now possible that the series of states $(q_0p_0)(q_1p_1)(q_2p_2)\ldots$ may not close. If it does close after m stages we have again $\lambda_l^m=1$, and if we write $\lambda_l=\exp i\sigma_l t$ there will be solutions that repeat themselves exactly after time $2\pi/\sigma_l$. If the system starts in state (q_0p_0) it will, in fact, be in some state (q_1p_1) at time t, since otherwise $\mathrm{K}=0$, and so on. There can be only one such path if all such sets close after time mt, for if there were two possible states at an earlier time we should be led to a contradiction of 1(18), just as the finite case led to a contradiction of 2(3). Hence each F_l will define a possible path with a definite period, and there is no possibility of transition from one such path to another. In the (q, p) space the path is along a locus such that $|\mathrm{F}_l(q, p)|$ has its greatest value. We are thus led to something like the steady states of quantum mechanics, though cyclic states would be a better description. It would be incorrect to speak of such paths as belonging to a particular eigenfunction, since if $F(q', p')$ is a point-concentration it will require an infinite series of orthogonal eigenfunctions to represent it; but all of them will be associated with sub-multiples of the fundamental period. It follows that for given $q_0 p_0$ a maximum of $|F_l(q'p')|$ is common to all of them, for K will vanish at all other points. If the series of states does not close in a finite number of stages for some choice of the interval t there will be no periodicity. The Darwin electron is evidently an instance.

We have not shown that every possible pair of values of (q, p) can be a stage of a periodic path even when all the eigenfunctions define cyclic paths. The eigenfunctions may in fact be an enumerable set, and so may be the points of maximum moduli for variations of p given q. The possible values of (q, p) are a continuous set and not all can be points of maximum modulus for some eigenfunction*. But a point concentration at any point can be represented by an expansion in terms of the eigenfunctions (with the aid of Euler summation or Stieltjes integration). It appears that the subsequent distribution can be regarded as expressing a definite probability for the system to be in any cyclic orbit, combined possibly with a distribution of the coefficients of the harmonics corresponding to each orbit such as will express a specified degree of knowledge about its initial position in that orbit. It seems possible that this will involve a discontinuity in $G(q'p')$ at time 0; such discontinuities are known to be possible in probability theory (Fréchet, p. 204). In other words, if (q, p) is not on a cyclic path, the system will immediately adopt some cyclic path, and there will be a definite distribution of probabilities among the paths possible.

4. We have reached a stage where comparison with wave mechanics shows some agreement and some serious differences. We can say that if the fundamental operation has discrete eigenfunctions strictly cyclic orbits of finite period will be possible, identifiable with the steady states of wave mechanics; but these may not include all the possible initial conditions. If the eigenfunctions are a continuous set the orbits may not be periodic and we can deal with free electrons. In classical mechanics they are necessarily a continuous set. But current wave mechanics shows no sign of the higher harmonics associated with each orbit. (They appear at the beginning of Heisenberg's treatment, but are modified beyond recognition later). There appears here to be an inconsistency, since $\psi\bar{\psi}$ for a steady state is interpreted in wave mechanics as the probability density for the co-ordinates. It is independent of the time and therefore in the present theory could only be interpreted as the time average of the probability density for a cyclic orbit. But such an orbit must be of

---

* Compare the function $f(x)=x(0 \leqslant x \leqslant 1)$; $f(x)=\frac{\pi-x}{\pi-1}(1 \leqslant x \leqslant \pi)$. This can be expanded in a sine-series for $0 \leqslant x \leqslant \pi$, but no term has a maximum at $x=1$.

finite size. Hence this average must be 0 for, say, all values of r for a hydrogen atom beyond a fixed value. The wave-mechanics solution, however, is finite at all distances, tending only to 0 at infinity. It seems therefore that $\psi\bar{\psi}$, when independent of time, is not correctly interpreted as a probability density for a steady state. It does tend to zero at infinity like $e^{-r/a}$, where a is a constant, and this fact suggests that if it is a probability density at all it may be the time-average of that for an electron that has been given an impulse in an unknown direction just strong enough to take it to infinity. A stronger disturbance would not give such an average since the probability density at all points would tend to 0 with time. The direction factors, however, do not appear easily capable of such an interpretation.

One constructive result that arises from this analysis is that in a steady state all the variables are periodic in time $2\pi/\sigma$, even, say, in atoms with many electrons. This in turn suggests a quantization of the orbits, since with an arbitrary choice of radii the periods would not in general be commensurable.

K will be symmetrical for pairs of similar particles, and all F_l are, or can be rendered, either symmetrical or antisymmetrical in such pairs. If any F_l with $\lambda_l \neq 1$ is symmetrical there will be sets of values (q, p) such that the co-ordinates and momenta of such particles are equal and $|F_l|$ is stationary for all variations. If $|F_l|$ has its greatest value at such points, there will be a possible orbit such that the two particles remain together for all time. But this contradicts the Coulomb law. We have, therefore, something like the exclusion principle: if there is any F_l with $\lambda_l \neq 1$, and symmetrical for similar particles, repelling according to the Coulomb law, then $|F_l|$ cannot take its greatest value when the q's and p's for such particles are equal. This result may be compared with that of Eddington, who obtains the Coulomb law as a consequence of the exclusion principle; this at least suggests the converse, that the exclusion principle may be a consequence of the Coulomb law. Clearly if all the F_l with $\lambda_l \neq 1$ are antisymmetrical, any state with the q's and p's for similar particles equal has $F_l = 0$, and therefore cannot form part of an orbit even instantaneously.

The discussion, however, has not led where it was expected to at the outset. In the wave mechanics solutions for unsteady states, ψ is the sum of a series of the form $\Sigma\psi_n e^{i\nu_n t}$, and $\psi\bar{\psi}$ is the sum of a term independent of the time and others of the form $\psi_m\bar{\psi}_n \exp i(\nu_m - \nu_n)t$. I expected that the exponential factors would be identifiable with the λ_l, the eigenvalues of the integral equation connecting probabilities at different times. This has turned out not to be possible. The λ_l must be identified with $\exp i\sigma_l t$, where $2\pi/\sigma_l$ is an orbital period, and each σ_l is associated with a series of higher harmonics of which wave mechanics shows no trace. No reason is found for identifying the σ_l with the Schrödinger ν_n. The latter are closely connected with the energy, and I think that the presentation would be made clearer if they were not put into time factors. Current wave mechanics appears to reject definite orbits but to associate the ν_m with a period, which must be the period of an orbit if it is one at all. The present treatment seems to be the result of a historical accident.

Whether we start from the hypothesis that all $|\lambda_l|=1$ or from the principles of relativity of time and reversibility, definite orbits are reinstated. I think that their rejection is the result of an ambiguity in the statement of the uncertainty principle. This shows that in an observation, if a co-ordinate is to be observed with standard error μ, the corresponding momentum is altered by something of the order h/μ. But we cannot say from this that the momentum had not a definite value before the observation and another after it. To try to represent it by a single continuous probability distribution is impossible ; given an observation there is one distribution for the probability before the observation and another after it, and the two cannot be combined, being on different data ; the time is a datum in each case and the two times are different. The demand in wave mechanics that description must be restricted to observed quantities is impossible to satisfy ; for if all co-ordinates in a system were observed it would be raised to a stellar temperature, and hardly any atom in it could be in its ground state. Thus we should be forbidden ever to say that most of the atoms in a system are in their ground states, and the whole of statistical mechanics would break down. The steady state is not observed as such but inferred ; I contend that it is a legitimate inference, and that it can be described only in terms of exact orbits.

5. The limitation of the present account to steady states does, however, fail to describe transition phenomena, and therefore is incomplete. The only hope here seems to be to modify the principle of reversibility and admit values of $|\lambda|<1$. We have seen that such values will arise if there is any possibility of random transition ; in the finite case, if any a_{ij} besides the dominant elements is $\neq 0$. Now random transition is generally regarded as a possibility in quantum theory ; the return of an excited atom to its ground state is one case, radioactive emission another. The latter is the more accessible to experimental test. If we try to maintain determinism in this case we shall have to say that the instants of α-particle emission from the same nucleus are determinate given one of them. Consider then the collection of a specimen of radium emanation in a few hours, its mean life being some days. If determinism is true we must say that we have selected a set of Ra nuclei whose rupture would have been predictable within a few hours from the times of rupture of their ancestors, and that the next ruptures will all occur within an interval equal to the time of collection. Thus the specimen of emanation would be inactive for a few days, and would then break up completely within a few hours. What actually happens is that it decays exponentially ; and this is possible only if determinism is relaxed to the extent of permitting a small probability of random transition out of a steady state.

If we consider a system capable of two cyclic states but with a small probability α of transition in time t from some stage of cycle A to a stage of cycle B, and a probability β of a reverse transition, it is easy to show that the eigenvalue $\lambda=1$ is unaltered, and the moduli of the others reduced by quantities of the orders of α and β. If the probability of transition within a period is small it is legitimate to continue to speak of the system as being in one of two cyclic states, the probabilities that it will be in one or the other approaching in a long time the limiting ratio β/α. Since the transitions may occur at different stages of the cycles, the probability within each cycle will approach uniformity with respect to the stages. This amounts to saying that even if the angular co-ordinates are exactly known at time 0, we shall approach a time when they are completely unknown. Thus the uncertainty relation reappears; we approach a state where the angular momenta are exactly stated for each cycle and the corresponding co-ordinates unknown. This tendency is therefore quite analogous to that in classical statistical mechanics, where the approach to a uniform probability distribution for any initial data is traceable to the unpredictability of collisions in detail; here it is traced to the unpredictability of transitions. But here we do not say that the conjugates cannot be spoken of as known within a given uncertainty; we say instead that if there were such knowledge we could still make no useful predictions about some conjugates beyond a certain time. The useful information will then be expressed by integrating the probability distribution over all values of the unpredictable variables.

The usefulness of this type of analysis clearly depends on the probability of transition from a cyclic state within a period being small. If we neglect all interactions we can form a set of eigenfunctions with the property of strict complex orthogonality $\int\int F_l\bar{F}_m dqdp=0$. But even with interaction we may still express the probability in terms of the same functions. They are no longer exact eigenfunctions when the transition probability is allowed for, but the orthogonal property persists. The change in the $|\lambda|$ can then be taken into account by a slow variation with time in the coefficients A_l instead of by altering the eigenfunctions. The method will fail on account of slow convergence if the transition probabilities are large.

The chief result of this argument is that the Schrödinger $\psi\bar{\psi}$ is not correctly interpreted as a probability density for the co-ordinates on any data. It is so interpreted by Heisenberg (1930, p. 148), who, however, states it as an additional assumption, and it does not appear to be necessary for most of the applications of the wave mechanics method, which concern the determination of energies in the steady states and probabilities of transitions between specified states. With regard to the latter, an element of randomness seems necessary to fit the experimental facts, but it is still not clear that the most economical way of stating it has yet been discovered.

Summary.

The identification of the $\psi\bar{\psi}$ of wave mechanics, when it involves the time, with a probability distribution, is found to be equivalent to a generalization in probability language of the principles of the relativity

of time and the reversibility of path. Either hypothesis leads to the result that strictly cyclic changes occur, and definite orbits are as much a consequence of quantum theory as of classical theory. Neither leaves any room for random transitions, which it is one of the main objects of quantum theory to treat. It is suggested that the difficulty may be avoided by a limitation of the scope of the principle of reversibility ; but it does not appear that even when this principle is applicable the probability density is correctly given by $\psi\bar{\psi}$.

References.

(1) A. J. Ayer, 'Language, Truth and Logic.' London : Gollancz (1936).
(2) C. G. Darwin, Proc. Roy. Soc. A, cxvii. pp. 258–293 (1927).
(3) M. Fréchet, Méthode des fonctions arbitraires : Theorie des évènemeuts en chaine. Borel's Traité du Calcul des Probabilitiés, t. 1, fasc. 3 (1938).
(4) W. Heisenberg, 'The Physical Principles of the Quantum Theory.' Chicago : University of Chicago Press (1930).
(5) H. Jeffreys, Proc. Roy. Soc. A, clx. pp. 337–47 (1937).
(6) H. Jeffreys, 'Theory of Probability.' Oxford : Clarendon Press (1937).

LXXXV. *Thermodynamical Notes.*

By Prof. W. Wilson, F.R.S.*

[Received September 21, 1942.]

Historical.

Traditional notions and methods persist tenaciously in physical science as in other human activities. Kelvin, as late as the year 1849, still believed in the conservation of heat (caloric) and that Joule's experiments only *seemed* to be in conflict with it. Carnot still dominates thermodynamics, not only because he imposed on it the sound principle named after him, but largely through his conception of a *reversible cycle* (1824). It appears that his work became known to Kelvin through reading Clapeyron's account of it, given in 1834 †, but not generally known till much later ‡. It is to the latter that we owe the familiar graphical representation of Carnot's reversible cycle, while his analytical methods undoubtedly greatly influenced and guided Clausius in his successful effort to reconcile Carnot's principle with the results of such experiments as those of Joule and Hirn *. Indeed, they are prominent in present-day text-books. It is fitting, therefore, and desirable that this short essay should begin with a brief reference to the salient features of the work of Carnot and Clapeyron—especially of the latter, since his work is now either unknown or only vaguely guessed and sometimes misrepresented. To Carnot himself we owe the law, now called the *second law* of thermodynamics, which appears in his work † as a theorem about the efficiency of reversible engines (Carnot's principle). This he deduced from two principles which he regarded as more fundamental. One of these, the impossibility of a *perpetuum mobile*, or, as we sometimes call it since the days of Wilhelm Ostwald, the *perpetuum mobile* of the first kind, is still accepted. The other one, the conservation of heat or caloric, is now rejected and was indeed rejected by Carnot himself some few years after the publication of his work ‡. His proof is described in A. W. Porter's little monograph §, and E. Mach has given a moderately complete and, on the whole, very fair account of the historical development of thermodynamics ‖, which includes the contributions of Carnot, Clapeyron, Clausius and the brothers Thomson. The familiar device of an infinitesimal reversible cycle, in which the work done is represented by an infinitesimal parallelogram, is due to Clapeyron. Though Mach's discussion of the merits and weaknesses of the outstanding contributions to thermodynamics is, generally speaking, very fair, he does convey a rather misleading impression of the part that conservation of heat played in Clapeyron's work. He quotes a passage in the latter's description of Carnot's cycle in which this principle appears to be used, but is not in fact used. The passage is to the effect that the isothermal compressional stage is prolonged till just as much heat has been given up to the sink as was removed from the source during the isothermal expansion. Lower down the same page of Mach's book we read : " Von dieser Vorstellung abgesehen, welche übrigens bei mehreren Entwicklungen durch besondere Umstände umschädlich wird, kann man die Clapeyron'sche Darstellung auch heute noch aufrecht halten." This amounts to saying that, apart from this view (conservation of heat) which in any case is rendered harmless (unschädlich) in many developments by special circumstances, we can still, even to-day, maintain Clapeyron's method (Darstellung). The fact is, however, that Clapeyron, while believing, doubtless, in the soundness of the conservation of heat, does not use it in deducing his formulæ. He uses the perfectly sound principle that the efficiency of the reversible engine is determined by the two temperatures. The fact that he thought of Carnot's principle as depending on the conservation of heat does not make his deductions from the principle unsound. It will be remembered that Newton described his space and time as *absolute* and the impression is conveyed, and there are some people who still believe that his mechanical theory is based to some extent on this attribute, while the fact is that his space and time differ only in a trivial way from those of Einstein's special theory of relativity. Newton's theory indeed emerges from the special theory when the constant, c, in the latter approaches the limit infinity.

---

* Communicated by the Author.

† E. Clapeyron, *Journal de l'Ecole Polytechnique*, vol. xiv. (1834).

‡ Clausius writes (Poggendorff's *Annalen*, 1850) that he was unable to find a copy of Carnot's work and that his knowledge of Carnot's theory was obtained from Clapeyron's paper.

---

* The problem of the reconciliation arose and became acute because of the belief that the conservation of heat was necessary for the validity of Carnot's principle.

† S. Carnot, ' Réflexions sur la Puissance Motrice du Feu et sur les Machines propres à développer cette Puissance ' (1824).

‡ A clear statement of the *first law* of Thermodynamics was found among Carnot's MSS. after his death, and also a computation of the mechanical equivalent of heat from the specific heats of air.

§ A. W. Porter, ' Thermodynamics,' Methuen's ' Monographs on Physical Subjects ' (1931).

‖ E. Mach, ' Principien der Wärmelehre ' (1900).

Clapeyron writes for the area of the infinitesimal parallelogram, bounded by isothermals and reversible adiabatics, the expression

$$(\partial p/\partial t)_v \, dt \, dv.$$

This he divides by dQ, the heat supplied. He then makes use of Carnot's principle, *i. e.*, he equates the expression he has obtained to

$$dt/\mathrm{C},$$

thus obtaining

$$dQ=\mathrm{C}(\partial p/\partial t)_v dv. \quad \text{. . . . .} \quad (1)$$

and the familiar

$$\mathrm{L}=(v_2-v_1)\mathrm{C}\, dp/dt\,^*. \quad \text{. . . . . . .} \quad (1\,a)$$

For Clapeyron C was an unknown function of the temperature ; but he and the two Thomsons were able to assign numerical values to it ; for various temperatures, in terms of the temperature scales of their times. James Thomson, for example, used Regnault's values for the latent heat of vaporisation of water and the corresponding values of dp/dt. He was thus enabled to make the famous prediction about the influence of pressure on the melting point of ice and one feels that, if any other name at all should be joined to that of Clapeyron in naming this equation, it should be that of Thomson rather than of Clausius. It was called Clapeyron's equation in my youth ; but for some reason or other the text-books, especially since the last war, insist on calling it the Clausius–Clapeyron equation, and some of them even venture to say that Clausius was the first to establish it rigorously. The fact is that his proof is not different from that of Clapeyron. Clausius did indeed take the important step of using the properties of an ideal gas to determine the form of the function C (Carnot's function, as it is called) †, and found it, naturally, to be proportional to the temperature, t, as defined by the equation $pv/\mathrm{R}=t$. But it was Kelvin who really appreciated the significance of the function and the legitimacy of using it to *define* a temperature scale. This he did in 1848 by (in effect) making C a constant. I prefer to call this scale the efficiency scale *, since its definition makes dt, the infinitesimal temperature difference between two isothermals, proportional to the efficiency of a Carnot engine working between them. Somewhat later Kelvin defined the well-known work scale of temperature by making C equal, or proportional, to the temperature. Clausius, notwithstanding the great intellectual achievement of reconciling Carnot's principle with the first law of thermodynamics †, does not seem to have quite grasped the significance of Carnot's function and of the nature of a temperature scale. On page 81 of 'The Mechanical Theory of Heat' (Browne's translation, 1879) we read : "Of all classes of bodies the perfect gases are best adapted for such a determination ‡, *since their laws are most accurately known* (the italics are mine). This seems to show that Clausius, at the time this was written, had some vague notion of a *true* temperature scale and that the temperature as it appears in $pv=Rt$ had some claim to be identified with this temperature. A. W. Porter would appear to have derived his conception of a temperature scale from this passage in the work of Clausius. He writes ('Monograph,' p. 21): "The necessity of comparing on every occasion with a standard substance can be avoided if once for all the function of temperatures which expresses the value of the efficiency is determined. *We cannot determine this from any theory.* It is necessary to select a substance *whose properties we have reason to believe are most easily ascertained*" (my italics). And so, like Clausius, he turns to the ideal gas. Even Planck § defines his temperature scale by the use of an ideal gas. The fact is, of course, as Kelvin long ago explained, that the definition of a scale of temperature is a conventional matter and can be effected without reference to any substance, simply by using Carnot's function.

---

* Clapeyron uses the letter k for latent heat and his formula refers to the unit volume of the vapour.

† Maxwell uses the term "Carnot's function" for 1/C : 'Theory of Heat,' p. 160 (edition of 1897).

Clapeyron arrived at a very interesting expression for the function C. He wrote dQ in the form

$$dQ=\frac{\partial Q}{\partial p_v}dp+\frac{\partial Q}{\partial v_p}dv,$$

and combined it with the isothermal condition

$$0=\frac{\partial t}{\partial p_v}dp+\frac{\partial t}{\partial v_p}dv,$$

thus obtaining for dQ

$$dQ=\left(\frac{\partial Q}{\partial v_p}-\frac{\partial Q}{\partial p_v}\frac{\partial t}{\partial v_p}\frac{\partial p}{\partial t_v}\right)dv.$$

---

* It may be noted that the zero of the efficiency scale is at minus infinity.

† See the observations made under the heading *Second Law.*

‡ He is referring to the determination of the form of the function $\phi(t_1, t_2)$ in $Q_1/Q_2=\phi(t_1, t_2)$. See equation (5).

§ M. Planck, 'Thermodynamik' (1905).

When this is substituted in (1) there emerges

$$C = \frac{\partial Q}{\partial v_p}\frac{\partial t}{\partial p_v} - \frac{\partial Q}{\partial p_v}\frac{\partial t}{\partial v_p}, \quad \ldots \ldots \quad (2)$$

a formula which is equivalent to

$$c_p - c_v = C\frac{\partial p}{\partial t_v}\frac{\partial v}{\partial t_p}, \quad \ldots \ldots \quad (2\,a)$$

as we see when we multiply both sides of (2) by $(\partial p/\partial t)_v \,.\, (\partial v/\partial t)_p$.

Reversibility.

In its traditional use (in thermodynamics) the term reversible is applied to any process, especially a cyclical process, every step of which can be exactly retraced. By "every step" is meant every change whatsoever that is involved in, or associated with the process. A reversible process, in this sense, is therefore such that, not only can the original state of affairs be *completely* restored, but this can be done by the exact retracement of every step in it. This is the reversibility that Carnot made use of in the deduction of his fundamental principle from the principle of conservation of caloric and the impossibility of the *perpetuum mobile* already referred to. It is also the reversibility that Clapeyron, Clausius and the Thomsons had in mind and which appears in most present-day works on thermodynamics. Only Planck, so far as I know, defines reversibility in a different way *. He begins by defining an *irreversible process* as one of such a kind that the original state of affairs cannot be restored everywhere *by any means whatsoever*. Any other process he calls *reversible*. Planck's definition of a reversible process therefore includes those processes which are reversible in the older and traditional sense and does not exclude others.

The condition of reversibility—in the traditional sense of the term—is insisted on in certain thermodynamical arguments and, more especially, in certain definitions. The most familiar type of thermodynamical system, one for which the equation of state is a relation between t, v and p (temperature, volume and pressure)—I shall call it a t, v, p system—will serve as an illustration. Thermodynamical arguments about such a system are indeed frequently carried out with the aid of Carnot's conception of a *reversible* engine, as for example, the deduction of Clapeyron's equation, which teachers and text-books mostly effect by Clapeyron's method. This method is valuable for introducing students to thermodynamics, but it is desirable to go beyond it and cut adrift from the artificiality of the reversible engine. For this purpose the usual procedure is to use the second law in the following form (Clausius) :— There is a function, S, whose value is completely determined by any two of the variables t, v and p, *i. e.* its differential is a, so-called, perfect or complete differential. This function is defined by

---

* M. Planck, *op. cit.* p. 80. The peculiar significance and implications of Planck's definition are discussed under *Second Law.*

$$dQ/f(t) = dS, \qquad (3)$$

where dQ is the heat communicated to the system *reversibly*—this is insisted on—$f(t)$ is a universal function of the temperature, which we may use to define a temperature scale [*], *e. g.* by making $f(t)$ equal to t (Kelvin).

Now it is *necessary* that a process shall consist in a succession of states of equilibrium if it is to be reversible in the traditional sense. It follows therefore that the work, dW, done *by the system* must, in a reversible change, be equal to $p\,dv$, where p is, evidently, the pressure represented by this letter in the equation of state, the variables in which, namely, t, v and p, expressly refer to a state of equilibrium [†]. In such a case, therefore,

$$dQ = dU + p\,dv,$$

where dU is the associated increment of the internal energy. Consequently

$$dS = (dU + p\,dv)/f(t). \qquad (3\,a)$$

When we do not make direct use of Carnot's engine it is invariably this statement (3 *a*) which forms our starting point. Sometimes we use it directly by applying the rule expressed by (6) below, or we may use it in building up or expressing other complete differentials, such as Gibbs' function for example. The following is a simple illustration:—We are given, shall we suppose, that the internal energy of a gas is completely determined by its temperature alone, *i. e.* $(\partial U/\partial v)_t = 0$, and also that c_v, its specific heat at constant volume, is a constant. We want to find out as much as we can about its equation of state. We begin by writing (3 *a*) in the form:

$$dS = \left(\frac{\partial U}{\partial t_v}\,dt + \frac{\partial U}{\partial v_t}\,dv + p\,dv\right) \Big/ f(t),$$

or, remembering what is given and adopting for simplicity Kelvin's work scale,

$$dS = (c_v dt + p\,dv)/t.$$

Since dS is a complete differential we may apply the rule

$$\frac{\partial}{\partial v}\left(\frac{c_v}{t}\right) = \frac{\partial}{\partial t}\left(\frac{p}{t}\right). \qquad (4)$$

Therefore $$0 = -\frac{p}{t^2} + \frac{1}{t}\frac{\partial p}{\partial t_v},$$

---

[*] It will easily be seen that

$$\frac{1}{f(t)}\frac{df(t)}{dt} = \frac{1}{C}.$$

[†] The work, dW, is *not in general* equal to $p\,dv$, where p is the pressure in the equation of state. When a piston is travelling outwards at a high speed in a cylinder the work done during the expansion dv is $p'dv$ where $p' < p$.

and so we find

$$\frac{p}{t}=\phi(v),$$

where ϕ is a function which cannot be determined with the data provided. Thus we find that the pressure of the gas, at constant volume, is proportional to the Kelvin, work-scale, temperature.

What part has reversibility played in this argument? If we adopt (3 *a*) to define S, as I think we are entitled to do, reversibility as such is not involved at all. Indeed, a succession of states of equilibrium does not even imply a reversible process, since they may conceivably be states of unstable equilibrium. The second law, in the form (3 *a*) can be applied to the unstable portion of a van der Waals (or similar) isothermal—a part of the isothermal which is sometimes rather hastily written off as fictitious—and we reach the conclusion that Maxwell's rule for the position of the horizontal line which determines the equilibrium pressure between a liquid and its vapour is a sound one. The condition of reversibility is sufficient, but not necessary. In thermodynamics (as applied to t, v, p systems for example) we are concerned to find relationships between t, v, p, the internal energy, U, the entropy S and various other functions which are characteristic of the system when it is in or near equilibrium, and with the way in which they change when the system passes (as in the Joule-Kelvin porous plug experiment for example) from an initial state of equilibrium to a final and possibly widely different state of equilibrium. Nor does it matter in the slightest how the system passes (whether reversibly or not) from one state to the other. Reversibility is associated with the following feature of the work done: $dW=pdv$. If we reverse the sign of dv the pressure, p, remains unchanged —except by an infinitesimal amount. Thus $p\,dv$ is reversed in sign. *Irreversible* work always has the form Fds, where F is a force such as that overcoming friction and ds is the associated displacement. It is characteristic of F that it *changes sign* when the sign of ds is reversed, so Fds is *not* changed in sign with ds. This could not happen if F were a variable in an equation of state. The Joule (or ri^2) development of heat in a circuit is of this nature, since in fact

$$ri^2dt=ri\times dq,$$

where dq is the quantity of electricity which has passed and ri changes sign with dq.

There are some phenomena, *e. g.*, thermo-electric phenomena, the treatment is frequently unnecessarily encumbered by elaborate provisions for ensuring reversibility. Porter writes:* "We cannot employ the second law unless this irreversible heat"—he is referring to the Joule development of heat in a thermo-circuit—"can be diminished until it is negligible. This can be done by introducing an opposing E.M.F. into the circuit so as to reduce the current till its square is negligible.

---

* A. W. Porter, *op. cit.* p. 52.

Thermal conduction along the wire is reduced by making the wires thin enough."

The following would appear to be a sound and simple way of dealing with the thermo-circuit :—For convenience suppose the temperature, t_0, of one junction fixed. Let t be the temperature of the other junction and let θ represent the temperatures from point to point on the wires between t_0 and t. Let π and π_0 be the electromotive forces (Peltier coefficients) at the junctions and $\sigma d\theta$ the Kelvin electromotive force in the portion of a wire with the terminal temperatures θ and $\theta+d\theta$. The equation of state is a relation involving π, t, σ, etc., and, we may add for completeness, q, the quantity of electricity that has passed. It is this equation of state we want to investigate. The variables in equations of state are measured, as we have seen, when the system is in equilibrium, so that Porter's requirement of a back E.M.F. is sound ; but its purpose is to achieve equilibrium, not reversibility. Then the wires must be long—it is immaterial whether they are thin or thick—so that t and θ shall have meanings ; not for the purpose of reducing the conduction of heat. Temperature, it will be remembered, is only defined for thermal equilibrium. The Clausius entropy equation, corresponding to (3 *a*) becomes for this case :

$$dS=\frac{1}{f(t)}\frac{\partial U}{\partial t_q}dt+\frac{1}{f(t)}\frac{\partial U}{\partial q_t}dq+\frac{1}{f(t)}\pi dq$$

$$-\frac{1}{f(t_0)}\pi_0 dq+dq\int_{t_0}^{t}\frac{\sigma_A-\sigma_B}{f(\theta)}d\theta,$$

or, introducing Kelvin's scale and applying the rule corresponding to (4),

$$\frac{\partial}{\partial q}\left(\frac{1}{t}\frac{\partial U}{\partial t_q}\right)=\frac{\partial}{\partial t}\left\{\frac{1}{t}\frac{\partial U}{\partial q_t}+\frac{\pi}{t}+\int_{t_0}^{t}\frac{\sigma_A-\sigma_B}{\theta}d\theta\right\},$$

$$\frac{1}{t}\frac{\partial^2 U}{\partial q\partial t}=\frac{1}{t}\frac{\partial^2 U}{\partial t\partial q}-\frac{1}{t^2}\frac{\partial U}{\partial q_t}+\frac{1}{t}\frac{\partial \pi}{\partial t_q}-\frac{\pi}{t^2}+\frac{\sigma_A-\sigma_B}{t}.$$

Now $(\partial U/\partial q)_t$ is zero. Therefore

$$0=\frac{1}{t}\frac{\partial \pi}{\partial t}-\frac{\pi}{t^2}+\frac{\sigma_A-\sigma_B}{t},$$

which may be written

$$\sigma_A-\sigma_B=\frac{d\pi}{dt}-\frac{\pi}{t},$$

since π is independent of q. We have reached this result without worrying about irreversible phenomena.

When the internal energy of a cell is a function of the temperature, t, and the quantity of electricity, q, which has passed, and when its E.M.F. is a function of t and q, or of t alone, we can deduce for it the familiar Gibbs–Helmholtz equation, even if the cell should happen to be, as it well might, an irreversible one.

In all this no mention has been made of the *direction* of a thermodynamical process and the formulæ of thermodynamics are in fact independent of directional criteria—as indeed is evident from the fact that they concern terminal states of equilibrium and do not refer to the process of transition from one to the other. Even the definition of entropy, whether we take the form (3) or (3 *a*), contains no reference to direction. Obviously it cannot. But Clausius, it will be remembered, emphasized that

> "DIE ENTROPY DER WELT STREBT EINEM MAXIMUM ZU."

This suggests some scrutiny of the second law.

Second Law.

The great achievement of Clausius was to show how the work of Joule, Hirn and others could be reconciled with Carnot's principle, which seemed to Kelvin and Joule to necessitate the conservation of heat. Joule even suggested * that Carnot's principle should be abandoned. In the hands of Clausius the principle becomes

$$Q_2/f(t_2)=Q_1/f(t_1), \quad \dots\dots\dots \quad (5)$$

which leads to the entropy definition (3) ; Q_2 and Q_1 being the quantities of heat communicated isothermally (some would add "and reversibly" but "isothermally" should imply that) to a thermodynamical system in passing from one reversible adiabatic to another at the temperatures t_2 and t_1 respectively. The function $f(t)$ is the universal function which we have met with in equation (3). It follows from (5) that the function S defined by (3) is completely determined by the values of the independent variables, *e. g.*, t and v, which fix the state of the system, and that it is independent of the process (reversible or otherwise) or of the path on a t, v diagram, by which the values of t and v have been reached. It is hardly necessary to point out that when ϕ is such a function of the independent variables x, y, z, etc.,

$$\left.\begin{aligned}\frac{\partial^2\phi}{\partial x\partial y}&=\frac{\partial^2\phi}{\partial y\partial x},\\ \frac{\partial^2\phi}{\partial x\partial z}&=\frac{\partial^2\phi}{\partial z\partial x},\end{aligned}\right\} \quad \dots\dots\dots \quad (6)$$

etc.

Such equations as (6), where the function ϕ may be the entropy, or any other function of the state of a thermodynamical system, provide the chief rule for obtaining thermodynamical results.

Clausius, Kelvin and Planck have felt it necessary to complete the theoretical edifice of thermodynamics by deducing the second law, as expressed analytically (by (3) or (3 *a*) for example) from some simpler statement which seems more obvious or axiomatic. Clausius chose the statement : "Heat cannot, of itself, pass from a colder to a hotter body." This is not a very good foundation, since the words "of itself" need a far too elaborate elucidation. Planck's axiom (Ostwald's *perpetuum mobile* of the second kind) is much superior. It may be stated as follows :—"It is impossible to construct a machine which performs a cycle of operations periodically, raises a weight, cools a reservoir and does nothing else." It is unnecessary, as Planck himself pointed out and explained, that the temperature of the reservoir should be lower than that of the surroundings. It is easy to deduce Carnot's principle and hence equation (5) and its consequences by the application of this axiom. Planck, however, did not use it in this way. Before proceeding to discuss Planck's attitude to the second law, it is worth while to point out that the *perpetuum mobile* of the second kind does not forbid the possibility of an engine which cools a reservoir and gives up some heat to another reservoir at a *higher* temperature. Let us inquire what the assumption of this latter possibility leads to. To begin with, it is very obvious that the axiom requires that *either* all engines, using two reservoirs, a source and a sink, must, in doing work, take heat from a source of heat at a higher temperature and reject some of it to a sink at a lower temperature, *or* all such engines must take heat from a source at a lower temperature and reject some of it to a sink at a higher temperature. Both possibilities cannot co-exist, for if *both* kinds of engines were assumed to be possible, we could take one of each sort, let them use the same two heat reservoirs, A and B, complete their cycles in the same time and in such a way that the heat taken from one of the reservoirs (shall we say from A) by one engine is exactly equal to that given up to it by the other engine. We could arrange that both engines contribute to the raising of the same weight. The first law requires that the heat abstracted from the other reservoir, B, by one of the engines must exceed that returned to it by the other, and, be it noted, one engine is not driving the other backwards. We thus have in the combination a machine of the kind which Planck's axiom declares to be impossible. Suppose then we were to make the assumption that all heat engines using two heat reservoirs are of the type which reject heat to a sink at a *higher* temperature than that of the source from which they draw heat. We can now apply Planck's axiom to show that all such engines (if they are reversible and work between the same two isothermals and any pair of adiabatics) are equally efficient. Thus we can deduce equation (5) ; *but*

---

* J. P. Joule, Phil. Mag. vol. xxvi. (1845).

$$f(t_2) > f(t_1)$$

when

$$t_1 > t_2,$$

A hypothetical Kelvin might adopt the convention

$$f(t) = -t,$$

or alternatively

$$f(t) = 1/t,$$

and the assumption we contemplated (rejection of heat to a sink at a higher temperature) could only become a reality if we were to alter the sign of t as usually defined, or to replace it by its reciprocal, or to adopt any other suitable convention which would invert the numerical order of the temperatures.

Now the second law, as I have presented it here, contains *no explicit reference to direction* ; but it happens that we can infer that *the algebraic sum of all entropy changes must be positive, with zero as a limit*, for processes consisting of a succession of equilibrium states. All that precedes in this paper has been deduced, or is deducible, without any knowledge of this fact about entropy. The directional criterion in classical thermodynamics *would appear* to be a *consequence* of the second law, rather than part of it. On the other hand, Planck and others find the significance of the second law to lie in the fact that it provides the criterion just mentioned. " Die Bedentung des zweiten Hauptsatzes der Thermodynamik besteht darin, dass er ein notwendiges und hinreichendes Kriterium dafür liefert, ob ein bestimmter in der Natur stattfindender Prozess reversibel oder irreversibel ist *." It is this view of the second law which led Planck to his definitions of irreversibility and reversibility and to apply his axiom (the *perpetuum mobile* of the second kind), not to a reversible engine, but to show that, when a gas expands adiabatically into a previously exhausted space, the original state of affairs cannot be completely restored. He is then able to show that the entropy of an *ideal gas*, as defined by $\int dQ/t$ (t being the ideal gas temperature—$t=pv/R$), cannot be reduced without leaving changes elsewhere, and finally, that the sum total of entropy cannot be reduced. I cannot help finding in this a slight blemish, not in the logical rigour, but in the logical beauty of Planck's thermodynamics. In his situation I should have felt it more satisfying to apply the impossibility of a *perpetuum mobile* of the second kind to show that the efficiency of all engines working between two temperatures t_1 and t_2 is equal to (in the limiting case) or less than $\phi(t_1, t_2)$, where ϕ is a universal function already referred to. The directional criterion is contained in the " less than." It must be contended, I think, that this criterion is consequential rather than fundamental in *classical* thermodynamics. It is only when we pass on to the statistical form of thermodynamics that we can succeed in giving Planck's view of the second law a satisfactory basis. Here the probability of the state of a system necessarily arises, and it happens that we can identify its logarithm (in the case of equilibrium) with the entropy of classical thermodynamics. We are thus enabled to widen the definition of entropy by making it the logarithm of a probability in all cases, and since we can maintain with confidence that changes (involving large numbers of elementary systems) must proceed in the direction of states of greater probability, so we can give the second law a more complete form, since we can associate the directional criterion with it when changes are actually taking place. Entropy in the statistical form of thermodynamics is defined for all states, whether equilibrium states or not. In the classical form of thermodynamics, on the other hand, entropy is only defined for equilibrium states.

---

* M. Planck, *op. cit.* p. 82.

LXXXVI. *Note on Dimensions.*

By Prof. W. Wilson, F.R.S.*

[Received September 21, 1942.]

Since correcting the proof of my short note on Dimensions I have had the opportunity of reading the proofs of Dr. Guggenheim's paper †.

His § 3 is not a *new* contribution to this discussion. It is contained in my paper published in the 'Philosophical Magazine' last January. The numerical quantities, which Dr. Guggenheim calls *definitional constants*, I discussed in a volume ‡ published nine years ago, where I spoke of them as being numbers associated with the definitions of units.

In § 7 I find again the old contention that, since *specific gravity* is the ratio of the density of a substance to that of water (or I presume to that of any other substance), it must be a pure number. But, clearly, were I to change the standard substance, say from water to mercury, the measured specific gravity would have its numerical value divided by 13·6 in consequence. What else can this mean except that specific gravity has dimensions? Indeed, it is just density, with the trivial difference that it is being expressed in terms of an arbitrary unit: that of water, or whatever it may be. The test of the dimensionless character of the ratio of two quantities is the *independence* of the numerical value of the ratio of that of either of the two quantities. Thus an angle, arc divided by radius, is independent of the value of the radius.

In § 9 I find an equation, numbered (9.3), namely

$$\frac{q_{A}\,q_{B}}{fr^{2}}=\epsilon\mathbf{a}. \qquad (1)$$

I shall use the letter K for the product $\epsilon\mathbf{a}$ and term it, as I think is usual, the dielectric constant of the medium in which the two charges q_{A} and q_{B} are situated. One of Dr. Guggenheim's expressions for the value of this constant is

$$\frac{1}{9\times10^{9}}\ \frac{\text{coulomb}^{2}}{\text{joule}\ .\ m},$$

where *m* obviously signifies a metre. Consequently

$$[K]=[\epsilon\mathbf{a}]=[Q]^{2}[F]^{-1}[L]^{-2}. \qquad (2)$$

Dr. Guggenheim chooses that **a** shall be a constant and that in empty space

$$K=\mathbf{a}. \qquad (3)$$

He then claims that it is obvious that $\epsilon\mathbf{a}$ has the same dimensions as **a** and that consequently ϵ is a pure number and that the situation is "free from ambiguity." It seems to me that the reason why ϵ is a number is that Dr. Guggenheim has *chosen* to assign the dimensions (2) to the other factor of K, namely, to **a**. The fact that ϵ is unity in one case does not surely remove its dimensions. All that we are *forced* to accept is that the quantity K, *i. e.* the *product* $\epsilon\mathbf{a}$, has these dimensions. When we express a physical quantity as a product of two or more factors we have an infinity of possibilities open to us, both as regards the numerical values of the factors and also their dimensions. One might endow ϵ with the dimensions of a length, whether or not it is unity in empty space, provided a suitable modification be made in the dimensions of **a**. Dr. Guggenheim might with more reason have assigned the dimensions (2) to ϵ; the other factor, **a**, would then have become one of the definitional constants he has spoken of in § 3. Had he done this I should have had no criticism to make.

---

* Communicated by the Author.

† E. A. Guggenheim, "Units and Dimensions," Phil. Mag. vol. xxxiii. p. 479 (1942).

‡ 'Theoretical Physics,' vol. ii. p. 36.

In § 13 I find a similar treatment of permeability. I shall use the symbol **M** for permeability to avoid confusion with Dr. Guggenheim's magnetic constant, for which he uses the symbol μ. We have

$$\mathbf{M}=\frac{\mu}{\mathbf{ac}^2}. \qquad (4)$$

This is how Dr. Guggenheim defines μ. His equations (13.1), for example, are

$$\mathrm{H}=2\frac{i}{r},\ \mathrm{B}=2\frac{\mu}{\mathbf{ac}^2}\ \frac{i}{r},$$

the latter of which I write, in the notation adopted for this note,

$$\mathrm{B}=2\mathbf{M}\frac{i}{r}.$$

The constant **a** in (4) is the same constant as that represented by this symbol in (1). The letter **c** denotes the velocity of electromagnetic waves in free space. It will be noticed that μ in (4) is a number because of the fact that **M** has been given the form of the product of three factors, namely, μ, $1/\mathbf{a}$ and $1/\mathbf{c}^2$, and that one of them, $1/\mathbf{a}$, has been given the dimensions of $1/\mathrm{K}$ and that the other has the dimensions of the reciprocal of the square of a velocity.

Obviously
$$u=\frac{1}{\sqrt{\mathrm{K}\mathbf{M}}}, \qquad (5)$$

where u is the phase velocity of electromagnetic waves in a medium of dielectric constant K and permeability **M**. We get, on substitution, Dr. Guggenheim's expression, namely

$$\frac{\mathbf{c}}{\sqrt{\epsilon\mu}}, \qquad (6)$$

but, as regards dimensions, all that is *forced* on us is

$$\left.\begin{aligned}\mathrm{LT}^{-1}&=\mathrm{K}^{-\frac{1}{2}}\mathbf{M}^{-\frac{1}{2}}\\ [\mathbf{c}/\sqrt{\epsilon\mu}]&=\mathrm{LT}^{-1}\end{aligned}\right\} \qquad (7)$$

and

together with (2).

The individual things **c**, ϵ, μ may, subject to (7) being satisfied, have any dimensions we please—not so K and **M**.

In conclusion, may I point out a trivial thing? The Maxwellian displacement, **D**, whether we use *ordinary* units or the Heaviside type, has usually been defined so that

$$\text{div}\ \mathbf{D}=\rho.$$

Maxwell, himself, wrote $\mathbf{D}=\mathrm{K}\mathbf{E}/4\pi$,

and gave other *equivalent* definitions. It has always, or mostly, in the past meant the quantity of electricity displaced through the unit area.

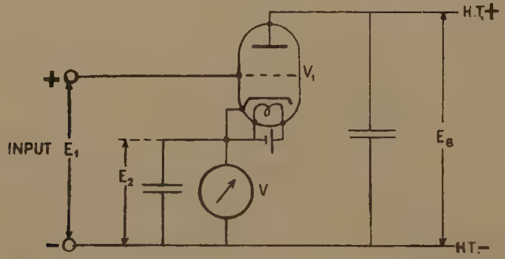
LXXXVII. *A Universal Valve Voltmeter with Linear Scale.*

By W. F. Lovering, B.Sc., University College, Nottingham *.

[Received September 15, 1942.]

Negative feedback can be used in a valve-voltmeter to reduce the influence of H.T. and L.T. voltages and valve characteristics on the calibration. The cathode-follower is an extreme form of negative feedback circuit in which the whole of the output voltage is fed back to the input. This circuit therefore lends itself to use in the production of a highly stable valve-voltmeter. The effect of negative feedback is to reduce sensitivity and to make the calibration curve approximately linear. The cathode-follower valve-voltmeter is thus less sensitive than the more usual types but its calibration curve is more nearly linear, and less dependent on circuit values. The normal valve-voltmeter is of little use for D.C. owing to the large effects of changes in H.T. or L.T. The cathode-follower avoids this trouble, and the circuit to be described is particularly suitable for D.C. work.

Fig. 1.



The cathode-follower circuit is modified, as in fig. 1, to act as a D.C. valve-voltmeter. V is a high resistance voltmeter of resistance R ohms; V_1 is a valve with high amplification factor (μ) and anode resistance R_a. The P.D. across the voltmeter, and therefore the voltmeter reading (E_2) will " follow " variation in E_1. Calculation shows that:

$$E_2 \doteqdot \frac{E_B+\mu E_1}{\dfrac{R_a}{R}+\mu+1}, \text{ (assuming straight characteristic curves),}$$

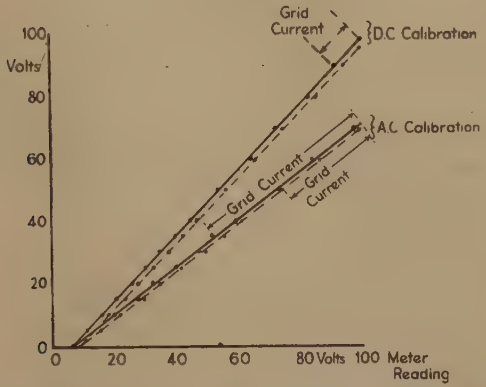
---

* Communicated by Professor L. F. Bates, D.Sc.

$$\doteqdot \frac{E_1}{1+1/\mu} + \text{a constant, if } E_B \text{ constant and } R \gg \frac{R_a}{\mu+1},$$

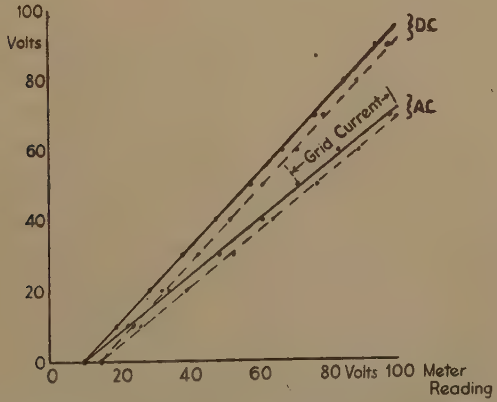
i. e., the calibration curve of the instrument is practically linear. Note also that for values of $E_1 \gg \frac{E_B}{\mu}$ change of E_B has very little effect on the

Fig. 2.



Calibration curves using valve type MH4 [μ=40, Ra=11,000 ohms].
Cathode voltmeter 100 v. (moving coil type), 1000 ohms per volt.
——— 120 v. H.T. supply. - - - - 240 v. H.T. supply.

Fig. 3.



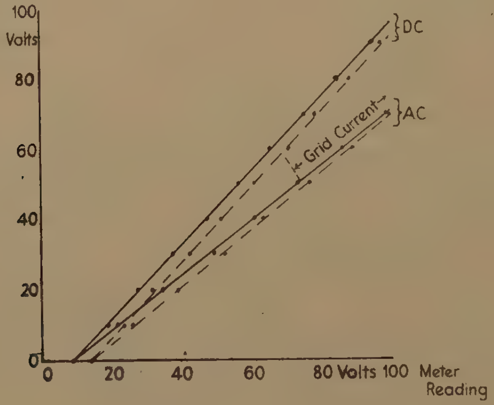
Calibration curves using valve type MHL4 (μ=20, Ra=8000 ohms).
Cathode voltmeter 100 v. (moving coil type), 1000 ohms per volt.
——— 120 v. H.T. supply. - - - - 240 v. H.T. supply.

calibration. (This is due to the negative feedback action of the resistance in the cathode lead.) Also the reading is largely independent of valve constants. The arrangement makes a convenient high impedance voltmeter for the range 5 to 250 V., for which electrostatic voltmeters are unobtainable.

If a condenser is used to by-pass the voltmeter in the cathode lead the same circuit can be used both as a D.C. valve-voltmeter and as a voltmeter giving peak readings on A.C. In the latter case, the steady bias developed by the D.C. anode current of the valve flowing through V keeps the valve biassed nearly to cut-off with no applied signal. When an A.C. signal is applied the valve acts as an anode bend detector, the resultant rise of I_a increasing the bias (and of course the meter reading) to a value just less than the peak valve of the A.C. signal. As in the D.C. case, the reading is almost independent of the H.T. voltage or valve characteristics except for small signals.

Grid current will flow when the grid potential falls to 1 V. to 2 V. negative with respect to the cathode. In a high resistance circuit this will render the meter reading useless. In practice grid current can be avoided by the use of a sufficiently high H.T. voltage—a voltage over twice the full scale reading of the D.C. voltmeter used is suggested.

Fig. 4.



240 v. H.T. supply.

Cathode voltmeter 100 v. (moving coil type), 1000 ohms per volt.

——— valve type MH4, $\mu=40$. – – – – valve type MHL4, $\mu=20$.

The arrangement is extremely flexible, and though best suited for maximum readings (*i. e.* full scale) of 20 to 250 V. can be arranged to read full scale on voltages as high as 1000 V. by use of the appropriate D.C. voltmeter. Owing to the danger of heater-to-cathode breakdown if an indirectly heated valve is used, one side of the heater should be connected to cathode and the filament transformer or battery should not be connected to any other circuit. Since fluctuation of H.T. and L.T. have little effect on the readings, a simple power pack can be used to operate the valve-voltmeter from a mains supply.

The calibration curves of a typical voltmeter are shown in fig. 2, which also shows the effect of a large change in the H.T. supply voltage. Figs. 3 and 4 show the effects of a radical change in amplification factor.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

LXXXVIII. *Some Notes on Newton's Chemical Philosophy written upon the Occasion of the Tercentenary of his Birth.*

Qui genus humanum ingenio superavit.

By Douglas McKie, D.Sc., Ph.D.,
Lecturer in the History of Science, University College, London *.

[Received October 15, 1942.]

Three Hundred Years ago, on Christmas Day of 1642, in the early months of the Civil War and almost between the camp-fires of the armies of King and Parliament, Isaac Newton was born in the manor-house of Woolsthorpe-by-Colsterworth, near Grantham, Lincolnshire, a posthumous child and hereditary lord of the manor. The scene was pleasantly and characteristically English. Here was Ermine Street, there the Great North Road ; here rose the grey stone tower of Colsterworth Church, and there ran the River Witham. But this spot, now destined to become famous for all time by the birth of a weak and puny infant, whose survival was greatly doubted, was already historic ; for, says Stukeley :

> 'This Mannor of Wulsthorp probably belong'd to Ulfus and took its name from him. He was 4th son of King Harold slain by William the Conqueror. His mother was Agatha, sister to Edwin and Morcar, earls, and sons of Algar, the great Mercian Duke, the remain of the royal blood of the Mercian kings ; who sprung from Stamford, and the Marsh country of Holland, whence the name of Mercian, and who had a very great estate in this country.
>
> 'Earl Morcar own'd Colsterworth, Skillington, Basingham, Stoke [Rochford] before mentiond, and Strawston, all in this neighbourhood. A wood near Horn Lane retains his name to this day, Morcar's wood ; near it Ulfsoaks, corruptly Walfox. King Harold, Ulfs father, had likewise a great estate in this country, both Stamford and Grantham. All that the Conqueror gave to his favourite, Robert de Todenei, Lord of Belvoir ; now the Duke of Rutland's. He ownd Uffington, Talington, Ropesly, Denton, Gretford, Braceborough, Sempringham, Greetham, Burley on the hill and many more townships. But Ulf, as being a son of Harold, was kept prisoner in Normandy by the Conqueror, together with all his relations, and Duncan son to the King of Scotland. He thought they might have some pretence to the Crown. King William Rufus set our Ulf at liberty, being a quiet honest man. He gave him his estate and honor'd him with knighthood ; and here at Wulsthorp in all probability he spent the remainder of his days, in great honor and contentment' (1).

So ran the story of Ulfsthorp or Wulsthorp or Woolsthorpe and of its neighbours near and far, 'Colsterworth, Skillington, Basingham, Stoke [Rochford] before mentiond, and Strawston, ... Uffington, Talington,

---

* Communicated by the Author.

Ropesly, Denton, Gretford, Braceborough, Sempringham, Greetham, Burley on the hill. . . .' And for this very English scene, whose village names sound like a medieval battle-roll, Newton, most understandably, never lost his affection. He always remained a countryman at heart; he loved to spend his time in Lincolnshire and among the people of his native county; he made his great discoveries at Woolsthorpe; he probably spoke with a broad Lincolnshire accent; and he had some intention of passing his days of retirement from active work in Grantham.

The details of Newton's education at the King's School, Grantham, and Trinity College, Cambridge, are well known to scientific readers, and so also are the main activities of his life and the record of his stupendous achievements in mechanics, physics and mathematics, namely, the formulation of the law of gravitational attraction, the discovery of the composition of light and the invention of the differential calculus. His chemical studies are, on the contrary, far from well known, although chemistry was one of his life-long interests. It seems therefore a fit moment to compile some notes on Newton's chemical studies and to attempt some appreciation of his chemical philosophy.

Early interests in chemistry: the letter to Aston (1669).

Newton's interest in chemistry probably began when he was lodging with Clark, the Grantham apothecary, while attending the King's School. He noted down interesting chemical recipes in his earliest note-books (2). He showed that he had a highly practical mind and great skill with his hands in those early days; he constructed mechanical toys and clocks and made sundials. Later on, in 1665 and in 1669, we have records that he bought chemical apparatus and materials, 'glass bubbles', *i. e.*, flasks, and 'glasses'; that he spent two pounds on 'aquafortis, sublimate, oyle pink, fine silver, antimony, vinegar, spirit of wine, white lead, salt of tartar, mercury'; that he bought a furnace for eight shillings and an 'air furnace' for seven shillings; and that he gave twenty-eight shillings for a copy of the 'Theatrum Chemicum' (3).

In this same year of 1669, Newton, who was soon to be elected to the Lucasian Chair of Mathematics, wrote on May 18th to his friend, Francis Aston, who was about to set out on a tour of the Continent and who had asked for some advice about his projected travels, and suggested, among other items for the traveller's attention, such as trades and arts, fortifications and navigation, diet and staple commodities, that he might:

> 'Observe the products of nature in several places, especially in mines, with the circumstances of mining and of extracting metals or minerals out of their oare, and of refining them; and if you meet with any transmutations out of their own species into another, (as out of iron into copper, out of any metall into quicksilver, out of one salt into another, or into an insipid body, &c.), those, above all, will be worth your noting, being the most luciferous, and many times lucriferous experiments too in philosophy . . . .
>
> 'As for particulars, these that follow are all that I can now think of, viz., Whether at Schemnitium, in Hungary, (where there are mines of

> gold, copper, iron, vitriol, antimony, &c.,) they change iron into copper by dissolving it in a vitriolate water, which they find in cavitys of rocks in the mines, and then melting the slimy solution in a strong fire, which in the cooling proves copper. The like is said to be done in other places, which I cannot now remember; perhaps, too, it may be done in Italy. For about twenty or thirty years agone there was a certain vitrioll came from thence, (called Roman vitrioll,) but of a nobler virtue than that which is now called by that name; which vitrioll is not now to be gotten, because, perhaps, they make a greater gain by some such trick as turning iron into copper with it, than by selling it. 2. Whether, in Hungary, Sclavonia, Bohemia, near the town Eila, or at the mountains of Bohemia near Silesia, there be rivers whose waters are impregnated with gold; perhaps, the gold being dissolved by some corrosive waters like *aqua regis*, and the solution carried along with the streame, that runs through the mines. And whether the practise of laying mercury in the rivers, till it be tinged with gold, and then straining the mercury through leather, that the gold may stay behind, be a secret yet, or openly practised . . .' (4).

A recent biographer of Newton concluded from this letter that its writer was 'a credulous alchemist', commented on his naive inquiries 'about the processes of changing iron into copper', and added that these inquiries 'give us a glimpse of his youthful ambition to devote himself to luciferous experiments, and of his dreams of turning such knowledge into the lucriferous creation of gold', while 'it was with great reluctance and after long years of labour that he finally became sceptical of the pretensions of alchemy' (5). For this verdict, there seems to be no evidence at all, as we shall presently show. As for the alleged conversion of iron into copper, it had only recently been shown by van Helmont that, contrary to the generally accepted opinion, the copper was already in the solution and was merely precipitated by the iron (6), an explanation that Newton accepted later on (see p. 862 below).

The contents of this letter to Aston and Newton's purchases of apparatus and materials show that he had much the same kind of interest in chemistry as his contemporaries had in that science and are no warrant for concluding that he was 'a credulous alchemist'. At this time too, Newton experimented on the production of alloys that would take a high polish and would therefore be suitable for use as mirrors in the reflecting telescope that he had recently designed (7); and it is said that he fitted up one of his rooms in Trinity College as a chemical laboratory and that he spent much time there in experimenting.

An æthereal medium perhaps the source of all things (1675).

In his famous 'Hypothesis explaining the Properties of Light', communicated to the Royal Society on December 9, 1675, Newton, after emphasizing that he did not assume this or any other hypothesis, but that he 'observed the heads of some great virtuosos to run much upon hypotheses', postulated for the sake of illustration a system arising from 'an æthereal medium' as follows:

> 'First, it is to be supposed therein, that there is an æthereal medium much of the same constitution with air, but far rarer, subtler, and more

> strongly elastic. Of the existence of this medium the motion of a pendulum in a glass exhausted of air almost as quickly as in the open air, is no inconsiderable argument. But it is not to be supposed, that this medium is one uniform matter, but compounded, partly of the main phlegmatic body of æther, partly of other various æthereal spirits, much after the manner, that air is compounded of the phlegmatic body of air intermixed with various vapours and exhalations: for the electric and magnetic effluvia, and gravitating principle, seem to argue such variety. Perhaps the whole frame of nature may be nothing but various contextures of some certain æthereal spirits, or vapours, condensed as it were by precipitation, much after the manner, that vapours are condensed into water, or exhalations into grosser substances, though not so easily condensible; and after condensation wrought into various forms: at first by the immediate hand of the Creator; and ever since by the power of nature; which, by virtue of the command, increase and multiply, became a complete imitator of the copies set her by the protoplast. Thus perhaps may all things be originated from æther' (8).

Newton, it may be noted, was here adopting Boyle's idea of the composition of air; as for the hypothesis that æther might perhaps be the origin of all things, he was presently to adopt another hypothesis and one that also had arisen in ancient Greece (9). But meanwhile, he went on to explain how terrestrial gravitational attraction might be regarded as caused by the condensation of an æthereal spirit, not of æther itself; this spirit bore the same sort of relation to æther as the 'vital aereal spirit', necessary for the maintenance of fire and life, bore to ordinary air. And 'condensation' went on concurrently with 'attenuation' in a continued cycle of change:

> '. . . . for nature is a perpetual worker, generating fluids out of solids, and solids out of fluids, fixed things out of volatile, and volatile out of fixed, subtil out of gross and gross out of subtil; some things to ascend, and make the upper terrestrial juices, rivers, and the atmosphere; and by consequence others to descend for a requital to the former' (10).

A possible mechanism of chemical change suggested in a letter to Boyle (1679).

In a letter to Boyle of February 28, 1678/9 (11), Newton showed that the suppositions he had made in the hypothesis concerning light outlined in the previous section—and he specially insisted that they were only mere suppositions—could be applied to explain various chemical phenomena. There was, he supposed, no abrupt change of density in the æther at the surface of a body, but a gradual decrease between the external æther and the internal æther across what he called 'the space of the æther's graduated rarity', within which space it is most important to note lay the superficies of the body. The chemical processes of solution, evaporation, ebullition or effervescence, and the 'transmuting' of 'gross' into 'aërial' bodies were then explicable in terms of a mechanism of repulsions and attractions dependent on this graduated rarity and its consequences. When bodies or particles approached one another, there was a rarefaction produced in the æther between them; and this rarefaction in turn produced in them an endeavour to recede from one another, because the æther strove to return to its former natural state of condensation. On the other hand, there was also produced a tendency to accede or adhere to one another on account of the pressure of the external æther surrounding the bodies, and when the bodies came so near together that the excess of pressure of the external æther above that of the interjacent æther was so great as to overcome their endeavour to recede from one another, they were violently driven together and adhered strongly. In such ways the repulsions and attractions in various chemical processes were explained.

But Newton had warned Boyle in the first paragraph of his letter :

> 'The truth is, my notions about things of this kind are so indigested, that I am not well satisfied myself in them ; and what I am not satisfied in, I can scarce esteem fit to be communicated to others ; especially in natural philosophy, where there is no end of fancying '.

Soon, Newton was to become less uncertain in speculations of this kind.

Newton's chemical experiments (*c. 1684–9*) : *Humphrey Newton's letters.*

We are now coming to the period in which Newton composed the immortal 'Principia'. Wearied with the burden of that supreme task, he turned to his other interest, the experimental science of chemistry, as a relaxation from the drudgery of mathematical labours. His amenuensis during these years, Humphrey Newton, stated in passages in two letters written in January and February 1727/8 to Conduitt, who was then collecting materials for a biography of Newton, that :

> '. . . . at spring and fall of the leaf, ... he used to employ about six weeks in his elaboratory, the fire scarcely going out either night or day, he sitting up one night and I another, till he had finished his chemical experiments, in the performances of which he was the most accurate, strict, exact. What his aim might be I was not able to penetrate into, but his pains, his diligence at these set times made me think he aimed at something beyond the reach of human art and industry ' (12).
>
> 'About 6 weeks at spring, and 6 at y^{e} fall, y^{e} fire in the elaboratory scarcely went out, which was well furnished with chymical materials as bodyes, receivers, heads, crucibles, &c., which was made very little use of, y^{e} crucibles excepted, in which he fused his metals ; he would sometimes, tho' very seldom, look into an old mouldy book w^{ch} lay in his elaboratory, I think it was titled *Agricola de Metallis*, the transmuting of metals being his chief design, for which purpose antimony was a great ingredient ' (13).

These two statements are partly contradictory ; and the second is gravely misleading. In the first, Humphrey tells us that he did not know what Newton was aiming at ; in the second, he says that his chief design was the transmutation of metals. So the witness contradicts himself. The second statement suggests that Newton consulted Agricola (his 'De re metallica ' (14) is meant) on the transmuting of metals. But Agricola would be one of the last authors to consult on such a problem ; that noble old metallurgist had a special dislike for alchemists and dowsers.

Then this second statement also suggests that Newton's use of antimony indicated that he was engaged upon the transmuting of metals. This is plain nonsense. Unfortunately, the substance antimony seems to have brought suspicion with it, as if it was Jezebel's own element, from the days of alchemy and the 'Triumphal Chariot' onwards, and evidently Humphrey thought that his master was touching pitch and being thereby defiled. He was possibly unaware that antimony had long been in use in alloys, for instance, those used in metal types for printing, for a considerable time. In fact, Newton had experimented from 1672 onwards on alloys for use as mirrors in his reflecting telescope (15), and he drew up an account of the composition of his specula, 'De metallo ad conficiendum speculum componendo et fundendo', the details of which show clearly that antimony was used as an ingredient (16). Humphrey's evidence is both unreliable and contradictory; it is also misleading and uninformed. His conclusion about Newton's consulting of Agricola has been swallowed by many writers without reference to Agricola's book and its author's opinions. Yet this is one of the main items of the evidence frequently adduced to prove that Newton was an alchemist!

Chemistry in the 'Principia' (1687).

With the publication of the 'Principia' in 1687 Newton's chemical philosophy became more precise. Already he had been greatly influenced by Boyle, and both he and Boyle had been close students and critics of Cartesianism. Yet, while Boyle had hesitated to adopt a thoroughgoing atomism, he had, for all that, been so successful in his atomistic explanations of chemical and physical phenomena that he might fairly be called the 'restorer of mechanical philosophy' so far as England is concerned. Newton's views, however, hardened to a true Demokritan atomism, probably as a result of his conclusion that the inter-planetary spaces must be void of all but the merest traces of matter if the solar system was not to run down. Newton's adherence to atomism is not explicit in the 'Principia': it begins to be evident in the paper 'De natura acidorum', to which we shall refer below, and it is set forth in all its clarity in the 'Queries' appended to the 'Opticks' in 1717.

Both Boyle and Newton were also influenced by the famous Netherlands chemist, Johan Baptist van Helmont (1577 or 1580–1644), whose posthumously published works were then having some considerable influence on the development of chemical thought (17). Van Helmont, who had discovered the materiality of gases and who had to some extent differentiated broadly between a number of gases, had revived the old Greek theory, the first of which we have any account, put forward by Thales of Miletus in the sixth century B.C. This theory regarded water as the source of all substances, air alone excepted. Form and state of aggregation were ever changing, but all things, except air, originated essentially from water. Newton, unlike Boyle, who was characteristically 'sceptical' about this theory, appears to have accepted it completely; and his adoption of it shows none of the hesitation with which he had so guardedly argued in 1675 and 1679 that all things might perhaps be produced from æther.

As for Newton's atomism, we shall presently see it become explicit, as the atomism of Boyle, or rather the atomism that Boyle was so skilfully employing in his physics and chemistry ; it was the old Greek atomic theory, which saw in the multifarious variety of substances in Nature merely so many aggregates of differing numbers and arrangements of ultimate identical atoms or ' uncuttables ' (ἄτομοι). By this theory, the atoms were all qualitatively the same ; it was their differing arrangements in shape and number that produced the variety of substances in the world ; and re-ordering in other numbers and re-arranging in other shapes would obviously lead to the formation of other substances, that is, to transmutation. It is in this sense that Newton may be considered a believer in the transmutability of matter ; and here he is at one with the modern atomic physicist. But this is not the transmutation and the alchemy of the Philosophers' Stone and the Elixir of Life !

Newton, therefore, took his chemistry from van Helmont and Boyle. The world he pictured in the ' Principia ' was a world, not of condensing and attenuating æther, but a world in which all substances were produced from water in ceaseless cyclic change, in a truly Herakleitian flux. A passage in the ' Principia ', where Newton is discussing the vapour emitted from the tails of comets, reveals this system to us in splendid detail :

> ' And it is not unlikely but that the vapour, thus continually rarefied and dilated, may be at last dissipated and scattered through the whole heavens, and by little and little be attracted towards the planets by its gravity, and mixed with their atmosphere ; for as the seas are absolutely necessary to the constitution of our earth, that from them, the sun, by its heat, may exhale a sufficient quantity of vapours, which, being gathered together into clouds, may drop down in rain, for watering of the earth, and for the production and nourishment of vegetables ; or, being condensed with cold on the tops of mountains (as some philosophers with reason judge), may run down in springs and rivers ; so for the conservation of the seas, and fluids of the planets, comets seem to be required, that, from their exhalations and vapours condensed, the wastes of the planetary fluids spent upon vegetation and putrefaction, and converted into dry earth, may be continually supplied and made up ; for all vegetables entirely derive their growth from fluids, and afterwards, in great measure, are turned into dry earth by putrefaction ; and a sort of slime is always found to settle at the bottom of putrefied fluids ; and hence it is that the bulk of the solid earth is continually increased ; and the fluids, if they are not supplied from without, must be in a continual decrease, and quite fail at last. I suspect, moreover, that it is chiefly from the comets that spirit comes, which is indeed the smallest but the most subtle and useful part of the air, and so much required to sustain the life of all things with us ' (18).

A further passage reads :

> ' The vapours which arise from the sun, the fixed stars, and the tails of the comets, may meet at last with, and fall into, the atmospheres of the planets by their gravity, and there be condensed and turned into

> water and humid spirits; and from thence, by a slow heat, pass gradually into the form of salts, and sulphurs, and tinctures, and mud, and clay, and sand, and stones, and coral, and other terrestrial substances ' (19).

Such was Newton's idea and understanding of ' transmutation '.

The concluding sentence of the first of these two passages throws an interesting light on Newton's recognition of the work of his contemporaries on the problem of combustion and respiration, the maintenance of fire and flame and life in our common air. He had already referred to the ' vital aereal spirit ' in his communication to the Royal Society on December 9, 1675 (see p. 850 above). Boyle, Hooke, Mayow and Lower had all studied this problem, and each of them had concluded in favour of something in the air, rather than the air itself, as the active agent supporting fire and life. Newton had evidently accepted their general conclusion as early as 1675. His acceptance of this and his mention of it are of so much chemical interest that it seems worth while on this occasion to quote the original from the first edition (1687) of the ' Principia ':

> ' Porrò suspicor spiritum illum, qui aeris nostri pars minima est sed subtilissima & optima, & ad rerum omnium vitam requiritur, ex Cometis præcipue venire ' (20).

The paper ' De natura acidorum ' (1692 and 1710).

The next item of chemical interest in Newton's publications is a paper entitled ' De natura acidorum '. This was probably written about 1692, and it is mainly concerned with the problem of chemical attraction. It has not, so far as we are aware, received any very serious attention hitherto. It was first published in 1710, in the supplementary volume to John Harris's ' Lexicon Technicum ' (London, 1704) (21). In the ' Introduction ' to this supplementary volume, it is stated that the paper was given by Newton to a friend in 1692; the text and a translation are given by Harris at the end of the ' Introduction ', which says that:

> ' The following Paper of Sir *Isaac Newton*'s is excellently well worth the Philosophical Reader's most serious and repeated Perusal; for it contains in it the Reason of the Ways and Manner of all Chymical Operations, and indeed of almost all the Physical Qualities, by which Natural Bodies, by their small Particles, act upon one another '.

This supplementary volume gives a list of subscribers at the end; and here ' Sir Isaac Newton, Knight ' is included for a ' Large Paper ' copy. Curiously, this book, containing Newton's only chemical paper, is not included in the list of books given by Colonel de Villamil as contained in Newton's library, which seems to suggest that there may possibly be other items missing from that list. Harris stated that he had Newton's permission to publish the paper; for, after speaking of ' that Incomparable Mathematician and Philosopher Sir Isaac Newton ', he says:

> ' But however, to do the Illustrious Author yet further Justice, I have, with his Leave, at the End of this *Introduction*, Printed a *Latin* Paper of his *De Acido*, with a Translation of my own; and which, tho' never

> Publish'd before, was given by him to a Friend, as long since as the Year 1692 . . . . .'.

The first part of the paper deals with the attractive force between the particles of acids and those of other bodies, and it seems to be a preliminary setting forth of ideas subsequently extended and developed in the 'Queries' appended to the 'Opticks'; the second part of the paper is entitled 'Cogitationes variæ ejusdem' and appears to be a series of disjointed observations and statements on the same problem and some extension of what had been set down in the first part. But the whole production is summary and brief and anticipatory of the 'Queries'. A note added at the end of the translation makes it clear that the text had been 'supervised and approved of by the illustrious Author'.

Generally, it may be said, Newton here pictured a world composed of different kinds of particles, such as acid particles, water particles, and so on, each kind of particle having its own characteristic attractive power. The particles also varied in size, and therefore in volatility. Solution was produced by the violent action of the acid particles on the particles of the body the acid was dissolving, their action being the result of attraction; they were so powerfully attracted in some cases that they shook asunder the particles of the body with such violence as to resolve them into air. Solution in water was produced by attraction between the acid particles and the particles of water. By the attraction of water, the particles of a dissolved salt receded from one another and were diffused throughout the liquid. A few quotations from Harris's translation will adequately illustrate the contents of the paper:

> 'The Particles of Acids are of a Size grosser than those of Water, and therefore less volatile; but much smaller than those of Earth, and therefore much less fix'd than they. They are endued with a great Attractive Force; in which Force their Activity consists; and thereby also they affect and stimulate the Organ of Taste, and dissolve such Bodies as they can come at. They are of a middle Nature between Water and Terrestrial Bodies, and attract the Particles of both.
>
> 'By this Attractive Force they get about the Particles of Bodies, whether they be of a metallick or stony Nature, and adhere to them most closely on all sides; so that they can scarce be separated from them by Distillation or Sublimation. When they are attracted and gather'd together about the Particles of Bodies, they raise, disjoyn, and shake them one from another; that is, they dissolve those Bodies.
>
> 'By their Attractive Force also, by which they rush towards the Particles of Bodies, they move the Fluid, and excite Heat; and they shake asunder some Particles, so much as to turn them into Air, and generate Bubbles: And this is the Reason of Dissolution, and all violent Fermentation . . . .
>
> 'The Particles of *Sal Alkali*, do consist of *Earthy* and *Acid* united together, after the same manner: But these Acids have so great an Attractive Force, that they can't be separated from the Salt by Fire; they do also precipitate the Particles of Metals dissolv'd in *Menstrua*, by attracting from them the Acid Particles, which before had dissolved them, and kept them suspended in the *Menstruum*.
>
> 'All Bodies have Particles which do mutually attract one another: The Summs of the least of which may be called Particles of the *first*

> *Composition*, and the Collections or Aggregates arising from the Primary Summs; or the Summs of these Summs may be call'd Particles of the *second Composition, &c.*
>
> 'Mercury and *Aqua Regis* can pervade those Pores of Gold or Tin, which lie between the Particles of *its last Composition*; but they can't get any further into it; for if any Menstruum could do that, or if the Particles of the first, or perhaps of the second Composition of Gold could be separated; that Metal might be made to become a Fluid, or at least more soft. And if Gold could be brought once to ferment and putrefie, it might be turn'd into any other Body whatsoever'.

This last paragraph is followed by a further statement that makes clear the attitude that Newton had adopted towards atomism:

> 'And so of Tin, or any other Bodies; as common Nourishment is turn'd into the Bodies of Animals and Vegetables'.

He had now concluded in favour of a world of Demokritan atoms and void; and thenceforth he is to be considered as accepting the old Greek atomic theory. If the 'first composition' of the ultimate (and identical) particles of any substance could be 'dissolved' or separated, then those particles might be re-grouped and re-arranged and re-ordered to form any other substance whatsoever. Such was the mechanism of the transmutability of all kinds of matter as pictured by Newton and also by most of his contemporaries.

The quotations we have given show that the ideas later set forth by Newton in the 'Queries' were considered in some detail by him as early as 1692. To those 'Queries' we shall now turn. But we wish to point out at this stage that many of these ideas were not new and that we are here concerned with their nature, not with their origins, although we shall, as a matter of interest, indicate those origins on occasion. Newton, it will be remembered, had in the second book of the 'Principia' rejected the Cartesian system of vortices as contrary to fact and inconsistent with the laws of motion; by his adoption of atomism, of a world of atoms and void, he had now also rejected the whole system of Descartes, the world of continuous matter and the plenum.

Chemistry in the 'Opticks' (1717).

The 'Opticks' first appeared in 1704 with sixteen 'Queries' appended to the text. In the second edition, which appeared in 1717 and was re-issued in 1718, the 'Queries' had increased to thirty-one, and the additions included the two essentially chemical 'Queries' numbered thirty and thirty-one. In these last two, we have the most complete account of Newton's chemical philosophy; but there are some details of chemical interest in the preceding ones, and we shall make brief reference to these before passing on to consider the contents of numbers thirty and thirty-one, which, it might be pointed out, take up practically as much of Newton's text as the twenty-nine others put together. The page-references given below refer to the fourth edition of 1730, substantially the same as the editions of 1717 and 1721.

The early queries deal with the interaction between matter and light. In Query 5 Newton asks :

> 'Do not Bodies and Light act mutually upon one another ; that is to say, Bodies upon Light in emitting, reflecting, refracting and inflecting [diffracting] it, and Light upon Bodies for heating them, and putting their parts into a vibrating motion wherein heat consists ? ' (p. 314).

His question in Query 7 suggests that the strength of the action between light and 'sulphureous Bodies' (*i. e.*, combustibles) is the cause why such bodies ignite more readily and burn more vigorously than other bodies do. Query 8 asks whether all 'fix'd Bodies' (*i. e.*, solids) do not emit light when heated beyond a certain degree and whether this emission is not produced by the vibrations of their parts ; and whether also all bodies that abound in 'terrestrial' parts, and especially 'sulphureous' (combustible) parts, do not emit light whenever their parts are sufficiently agitated either by heat or friction or percussion or putrefaction or any vital motion or other cause. Query 9 asks :

> 'Is not Fire a Body heated so hot as to emit Light copiously ? For what else is a red hot Iron than Fire ? And what else is a burning Coal than red hot Wood ? ' (p. 316).

From this there naturally follows Query 10, which asks :

> 'Is not Flame a Vapour, Fume or Exhalation heated red hot, that is, so hot as to shine ? For Bodies do not flame [inflame] without emitting a copious Fume, and this Fume burns in the Flame' (p. 316).

In fact, many combustibles emit a fume, when sufficiently heated, and this fume can be ignited. This can be seen in the distilling of spirits ; for, if the head of the still be removed, the vapour rising from the still can be ignited with the flame of a candle and the flame can be seen running along the vapour from the candle to the still. Moreover, if the flame of certain bodies, such as oil, tallow, wax, wood, coal, pitch, sulphur, is put out, the thick smoke, which, by burning, had constituted the flame, can then be seen. Flame is thus 'red hot Smoke'.

When gunpowder is fired, 'it goes away into flaming Smoke'. The charcoal and the sulphur in it are ignited easily and then they set fire to the nitre, thereby rarefying its 'spirit' into vapour, which rushes out explosively. The sulphur is also vaporized and this augments the explosion ; the acid vapour from the sulphur attacks the nitre violently and releases the 'spirit' of the nitre, so causing a 'great Fermentation', further increasing the heat and rarefying the 'fix'd Body' of the nitre into 'fume', and so the explosion is thereby made still 'more vehement and quick'. For, if salt of tartar (potassium carbonate) is mixed with gunpowder and the mixture fired by warming, the explosion increases in violence because the 'vapour' of the gunpowder rarefies the salt of tartar. So, according to Newton, the explosion of gunpowder is, in effect, a sudden vaporization of the powder produced by the violent heat generated by the interaction of the ingredients.

It is interesting to note in passing that in Query 28, when Newton rejects the notion of light as caused by 'Pression or Motion, propagated through a fluid Medium', after a detailed argument against such a medium, he points out that his rejection has the authority of the ancient atomists :

> 'And for rejecting such a Medium, we have the Authority of those the oldest and most celebrated Philosophers of *Greece* and *Phœnicia*, who made a *Vacuum*, and Atoms, and the Gravity of Atoms, the first Principles of their Philosophy; tacitly attributing Gravity to some other Cause than dense Matter' (pp. 343–4).

Turning now to Query 30, the first of the two specially concerned with chemistry, we see what Newton's idea of transmutation was and in what sense he understood the term. His query opens as follows :

> 'Are not gross Bodies and Light convertible into one another, and may not Bodies receive much of their Activity from the Particles of Light which enter their Composition ?' (p. 349).

At once, Newton points out again that all solid bodies, when heated hot enough, emit light and continue to do so as long as they are sufficiently hot. He adds : 'I know no Body less apt to shine than Water'; and yet water could, by repeated distillation, be changed into earth, and this earth, if heated sufficiently, could be made to 'shine', *i. e.*, to emit light, as other bodies did.

'The changing of Bodies into Light, and Light into Bodies', said Newton, 'is very conformable to the Course of Nature, which seems delighted with Transmutations' (p. 349). This last phrase led Newton's most recent biographer to state that 'Query 30 is a veiled apology for the search for the transmutation of metals' (22), which is a most remarkable conclusion. The same author goes on to say of Newton : 'He pictures Nature as delighted with transmutations, and gives examples to show that all bodies are composed of the same hard particles which differ in nature only because of their arrangement and their mutual forces of attraction. He even carried his belief in transmutability so far as to query whether gross bodies and light are not convertible into one another' (23). This is very ill-informed criticism; it lays the emphasis in the wrong place; there was nothing alchemical in such beliefs, which were currently accepted by natural philosophers in general. But this author, presumably on the same grounds, regarded Boyle also and even Hooke as alchemists!

Newton himself explains his ideas on these matters by giving a very useful illustration of what he meant by Nature's delight in transmutations in the case of water which might assume the forms of 'a very fluid tasteless Salt', a vapour or air, and ice or stone :

> 'Water, which is a very fluid tasteless Salt, she [Nature] changes by Heat into Vapour, which is a sort of Air, and by Cold into Ice, which is a hard, pellucid, brittle, fusible Stone; and this Stone returns into Water by Heat, and Vapour returns into Water by Cold' (p. 349).

In similar ways :

> 'Earth by Heat becomes Fire, and by Cold returns into Earth. Dense Bodies by Fermentation rarify into several sorts of Air, and this Air by

Fermentation, and sometimes without it, returns into dense Bodies' (p. 349).

As a further example, mercury is cited. This substance appears in various forms as either a fluid metal, or a hard brittle metal, or a corrosive salt (sublimate), or a volatile white earth (calomel), or a red earth (cinnabar), or a red or white precipitate, or a fluid salt, or, by distillation, a vapour, or even 'Fire', when agitated in a vacuum. 'And', says Newton, 'after all these Changes it returns again into its first form of Mercury' (p. 350). Such was Newton's view of transmutation, and he concluded his account of it with a general reference to the theory that he had adopted from van Helmont:

> 'Eggs grow from insensible Magnitudes, and change into Animals; Tadpoles into Frogs; and Worms into Flies. All Birds, Beasts and Fishes, Insects, Trees, and other Vegetables, with their several Parts, grow out of Water and watry Tinctures and Salts, and by Putrefaction return again into watry Substances. And Water standing a few Days in the open Air, yields a Tincture, which (like that of Malt) by standing longer yields a Sediment and a Spirit, but before Putrefaction is fit Nourishment for Animals and Vegetables' (p. 350).

Newton ended with the surely unobjectionable question:

> 'And among such various and strange Transmutations, why may not Nature change Bodies into Light, and Light into Bodies?'.

We may remind ourselves that in 1789 Lavoisier included light amongst the chemical elements; the phlogiston theory took it into particular account (24); and it has a continuous history in chemical thought up to our own time, a period in which Newton's views, far from seeming alchemical, are not at all unacceptable. It is curious that a modern biographer should have found anything to condemn as credulous in Newton's reasonable inquiry. The Query includes, as quoted above, Newton's adoption of Boyle's idea as to the production of 'Air', indeed of 'several sorts of Air' by the 'Fermentation' of 'dense Bodies'. Such a 'transmutation', like many other chemical changes, Newton pictured as produced by the re-arrangement of ultimate particles, qualitatively identical; yet this simple, but philosophically sound, mechanism of our chemical childhood, with its close correspondence to the findings of modern atomic physics, is still dubbed 'alchemy' by some authors. The same might be said of that other scientific 'fact', namely, the convertibility of water into earth, which was not refuted until Lavoisier announced the result of his classical researches in 1770. It is interesting to notice in passing that both these topics were mentioned in the paper 'De natura acidorum' to which we have already referred, and both are implicit in the quotation we have made from the 'Principia'.

In Query 31, the last and by far the longest since it takes up more than thirty pages of text, Newton set out his interesting views on what we now call, with a total disregard of the meaning of words, 'chemical affinity'. He opens with the question:

> 'Have not the small Particles of Bodies certain Powers, Virtues, or Forces, by which they act at a distance, not only upon the Rays of Light

> for reflecting, refracting, and inflecting [diffracting] them, but also upon one another for producing a great Part of the Phænomena of Nature ?' (p. 350).

The physical evidence for such a supposition is then made clear :

> 'For it's well known, that Bodies act one upon another by the Attractions of Gravity, Magnetism, and Electricity ; and these Instances shew the Tenor and Course of Nature, and make it not improbable but that there may be more attractive Powers than these. For Nature is very consonant and conformable to her self' (pp. 350–1).

Before proceeding to details, Newton points out that he does not propose to consider here how these attractions may be performed ; that attraction, as he understands the term, may be performed either by impulse or by some unknown means ; and that the term signifies any force by which bodies tend towards one another, irrespective of the cause of that force :

> 'For we must learn from the Phænomena of Nature what Bodies attract one another, and what are the Laws and Properties of the Attraction, before we enquire the Cause by which the Attraction is perform'd' (p. 351).

Gravitational, electric and magnetic attractions reach to very considerable distances and are commonly observed : there may be other attractions that extend to such short distances as to have hitherto escaped observation.

Attraction and the attractive powers of particles being thus admitted as 'not improbable', on the grounds of uniformity in Nature, Newton proceeded to consider in detail instances drawn from well-known chemical facts. The deliquescence of 'salt of tartar' (potassium carbonate), he ascribed to an attraction between its particles and the particles of water-vapour present in the atmosphere. On the other hand, common salt and salt-petre and vitriol did not deliquesce for lack of such an attraction. Moreover, a given amount of 'salt of tartar' would draw only a proportionate amount of water-vapour out of the air ; and here clearly its attractive force had been 'satiated' and it could take up no more water. The existence of this attractive power was also evident from the fact that 'a great Heat' was necessary to distil water from 'salt of tartar', whereas 'a gentle luke-warm Heat' was adequate to distil water alone.

There was a similar attraction, 'a like attractive Power', to be inferred between the particles of oil of vitriol and those of water, from very similar data, namely, the deliquescence of oil of vitriol exposed to the atmosphere, the limit to the proportion of water-vapour so taken up by the oil of vitriol, and the difficulty of taking away this water in distillation. Further, when these liquors were mixed, and when they grew very hot in the mixing, the development of heat surely argued a great motion in their parts ; and, in turn, this motion argued that, in mixing, their parts coalesced violently and rushed towards one another with an accelerated motion. In like manner, the heat and ebullition produced by the solution of iron filings in *aqua fortis* or in spirit of vitriol pointed to a violent motion of the parts, and that motion indicated in turn that the acid parts of the liquor rushed violently towards the parts of the metal, ran forcibly into its pores and so disrupted the particles from the mass of the metal. Again, the acid particles distil alone with a gentle heat, but they will not separate from the particles of the metal without a very violent heat; and so, asks Newton, 'Does not this confirm the Attraction between them?' (p. 353).

The production of hydrochloric and nitric acids by the action of sulphuric acid on common salt and salt-petre was explained as follows:

> 'When Spirit of Vitriol poured upon common Salt or Salt-petre makes an Ebullition with the Salt, and unites with it, and in Distillation the Spirit of the common Salt or Salt-petre comes over much easier than it would do before, and the acid part of the Spirit of Vitriol stays behind; does not this argue that the fix'd Alcaly of the Salt attracts the acid Spirit of the Vitriol more strongly than its own Spirit, and not being able to hold them both, lets go its own?' (p. 353).

Reactions of this kind could be very violent, as, for example, in the action of a mixture of oil of vitriol and spirit of nitre on spirit of wine—a reaction which seems to have provided the organic chemist of our time with something akin to the universal solvent. It might be said:

> 'Where the Dissolution is slow, it makes a slow Ebullition and a gentle Heat; and where it is quicker, it makes a greater Ebullition with more heat; and where it is done at once, the Ebullition is contracted into a sudden Blast or violent Explosion, with a heat equal to that of Fire and Flame' (p. 354).

And so the vigour of chemical change ranged from slow ebullition and gentle heat to violent explosion and the heat of fire and flame, and clearly the attractive forces varied as widely.

Here Newton produced another example, which he had taken, together with its application, from Lemery (25). The swelling and the inflaming of a mixture of sulphur and iron filings made into a paste with water was adduced to explain the cause of volcanic eruptions, earthquakes, etc., since there were adequate supplies of sulphur and minerals in the earth to reproduce these effects on a large scale. Further, thunder and lightning and fiery meteors were produced by the fermentation of 'sulphureous [combustible] Steams' ascending into the air and there fermenting with 'nitrous Acids':

> 'For the Air abounds with acid Vapours fit to promote Fermentations, as appears by the rusting of Iron and Copper in it, the kindling of Fire by blowing, and the beating of the Heart by means of Respiration. Now the above-mention'd Motions are so great and violent as to shew that in Fermentations the Particles of Bodies which almost rest, are put into new Motions by a very potent Principle, which acts upon them only when they approach one another, and causes them to meet and clash with great violence, and grow hot with the motion, and dash one another into pieces, and vanish into Air, and Vapour, and Flame' (p. 355).

The action between 'Salt of Tartar *per deliquium*' (the liquid produced by the deliquescence of potassium carbonate) and metallic solutions, whereby the metal was precipitated, showed that the acid particles were attracted more strongly by the ‘ salt of tartar ’ than by the metal and so passed from the metal to the ‘ salt of tartar ’. In similar ways a solution of iron in *aqua fortis* dissolved *lapis calaminaris* (natural zinc carbonate) and let go the iron, a solution of copper dissolved iron immersed in it and let go the copper (this does not look like the ‘ alchemy ’ that some authors have seen in the letter to Aston, quoted on pp. 848–9 above), a solution of silver dissolved copper and let go the silver, a solution of mercury in *aqua fortis* dissolved iron, copper, tin, or lead, and let go the mercury. Here it seemed to Newton that the acid particles of the solvent were attracted by the particles of the metal that it took up more strongly than by the particles of the metal that it let go, so that they were attracted more strongly by *lapis calaminaris* than by iron, more strongly by iron than by copper, more strongly by copper than by silver, and more strongly by iron, copper, tin, and lead than by mercury. And Newton went on to infer that it was, for the same reason, that iron required more *aqua fortis* to dissolve it than did copper, and that copper required more of the acid than the other metals did, and that this same reason also explained why iron, of all the metals, was dissolved most easily and was most apt to rust, and next to iron, copper. Here Newton was unknowingly brought up against a difficulty that was to baffle chemists for another century : the amounts of acid necessary to dissolve a fixed weight of these different metals depended also on something other than attraction, on a fundamental not yet revealed to chemical philosophers. For, while ‘ attraction ’ or ‘ affinity ’, as it is now misnamed, was, of course, a factor in these changes, the combining weight or equivalent of each, itself a function of atomic weight, the fundamental chemical unit, lay behind the data upon which Newton was turning the intrepid speculative power of his fertile genius.

Further illustrations of various kinds were then drawn from examples of different chemical changes : the water in acids is set free when these acids are attracted more powerfully by other bodies, such as metals ; mutual attraction caused the particles of one substance to be vaporized or sublimed by the particles of another ; the sulphur in cinnabar, when the latter was sublimed from quicklime, by the stronger attraction of the lime, let go the mercury and passed over to the lime ; and, in general, the ingredients composing substances appeared to unite by their mutual attractions.

Particles of a salt dissolved in a large amount of water do not sink to the bottom, although they are heavier than the water, but diffuse evenly through the water. Does not this imply that they recede from one another, endeavouring to expand and to get as far apart as the quantity of water will allow ? And so also, does it not imply that they have a repulsive force by which they move away from one another or at least that they attract the water more strongly than they attract one another ? Such was Newton’s idea of solution : and he added :

> ‘ For as all things ascend in Water which are less attracted than Water, by the gravitating Power of the Earth ; so all the Particles of Salt which

> float in Water, and are less attracted than Water by any one Particle of Salt, must recede from that Particle, and give way to the more attracted Water' (p. 363).

When such a solution is evaporated, its particles concrete in regular figures, which, Newton held, argued that before this concretion they floated in the liquid 'at equal distances in rank and file, and by consequence that they acted upon one another by some Power which at equal distances is equal, at unequal distances unequal'; for such a power would lead to an orderly arrangement of the particles, and without such a power they would be disordered and would concrete irregularly.

Newton proceeded to extend this, by means of the example of the refraction in Iceland spar, to its proper conclusion of polarity in the particles :

> 'And since the Particles of Island-Crystal act all the same way upon the Rays of Light for causing the unusual Refraction, may it not be supposed that in the Formation of this Crystal, the Particles not only ranged themselves in rank and file for concreting in regular Figures, but also by some kind of polar Virtue turned their homogeneal Sides the same way?' (p. 363).

Various attempts had been made to explain the powerful cohesion of the parts of all homogeneous solids. Some had invented hooked atoms, which, said Newton, was but begging the question: 'and others tell us that Bodies are glued together by rest, that is, by an occult Quality, or rather by nothing; and others, that they stick together by conspiring Motions, that is, by relative rest amongst themselves'. But Newton would have none either of the hooked atoms of Lucretius or the more modern Lemery or of the occult quality of Cartesian rest or relative rest :

> 'I had rather infer from their Cohesion', he said, 'that their Particles attract one another by some Force, which in immediate Contact is exceeding strong, at small distances performs the chymical Operations above-mention'd, and reaches not far from the Particles with any sensible Effect' (p. 364).

That all bodies seem to be composed of hard particles was evident from the congelation of fluids and the production of solid products by the reaction of fluids (*e. g.*, spirit of nitre and mercury, and spirit of urine and spirit of salt) :

> 'And therefore Hardness may be reckon'd the Property of all uncompounded Matter. At least, this seems to be as evident as the universal Impenetrability of Matter. For all Bodies, so far as Experience reaches, are either hard, or may be harden'd; and we have no other Evidence of universal Impenetrability, besides a large Experience without an experimental Exception' (pp. 364–5).

But compound bodies are also porous, besides being hard, some of them being very hard, and they consist of parts that are merely laid together : and so the component particles, which are, of course, not porous, must be very hard. And the particles of a mass can touch only in a few points, and therefore much less force is needed to separate them than would be necessary to break a solid particle itself:

> 'And how such very hard Particles which are only laid together and touch only in a few Points, can stick together, and that so firmly as they do, without the assistance of something which causes them to be attracted or press'd towards one another, is very difficult to conceive' (p. 365).

Newton then pointed to further instances of attraction between particles drawn from the phenomena we now refer to as those of surface tension, capillary attraction, the wetting of solids by liquids, the spreading of surface films, and so on, and he concluded:

> 'There are therefore Agents in Nature able to make the Particles of Bodies stick together by very strong Attractions. And it is the Business of experimental Philosophy to find them out' (p. 369).

The smallest particles may cohere by the strongest attractions, giving larger particles which had less powerful attractions; and these in turn may cohere and compose still larger particles with still less powerful attractions, and so on through many stages, until successive aggregations reached the largest particles upon which chemical phenomena depended and which by their cohesion produced particles of sensible magnitude.

The production of air and vapour seemed also to argue the existence of a repulsive force: as in algebra, where affirmative quantities vanish, there negative ones begin, so in mechanics, where attraction ceases, there repulsion should succeed. The attractive force of the particles extended only to short distances from the particles; but when heat or fermentation shook the particles from the bodies that they composed, they underwent a vast expansion, when once they were beyond the reach of the attractive force: they receded from the body that they had composed and from one another with great strength and sometimes took up more than a million times more space than they had occupied when concreted in the solid body. Now Boyle, whom Newton does not name in this connexion, had pictured the particles of air, in his experiments on vacuum with the air-pump, as like little fleeces of wool which expanded and contracted as the pressure varied. Newton disposes of this in favour of expansion by repulsion:

> 'Which vast Contraction and Expansion seems unintelligible, by feigning the Particles of Air to be springy and ramous, or rolled up like Hoops, or by any other means than a repulsive Power' (p. 371).

Whence it followed that the particles of fluids that do not cohere too strongly are readily vaporized; but the particles of denser bodies, bodies referred to as 'fix'd', *i. e.*, solids, are not separated without greater heat or perhaps a fermentation, and when they are rarefied, they become 'true permanent Air', those particles receding from one another with the greatest force and being brought together with the greatest difficulty that on contact cohere most strongly.

> ' And thus Nature ', adds Newton, ' will be very conformable to her self and very simple, performing all the great Motions of the heavenly Bodies by the Attraction of Gravity which intercedes those Bodies, and almost all the small ones of their Particles by some other attractive and repelling Powers which intercede the Particles ' (p. 372).

The *vis inertiæ* is merely a passive principle, Newton then pointed out, and of itself it cannot have produced the motion of the world. Some other principle is necessary for putting bodies into motion, and also a further principle is then necessary for the conserving of the motion. The Cartesian vortices, which Newton had already shown in Book II of the ' Principia ' to be inconsistent with known facts, were also insufficient to maintain the motion of the world. This motion must be both conserved and recruited by active principles, such as were the causes of gravity and of fermentation :

> ' For we meet with very little Motion in the World, besides what is owing to these active Principles. And if it were not for these Principles, the Bodies of the Earth, Planets, Comets, Sun, and all things in them, would grow cold and freeze, and become inactive Masses ; and all Putrefaction, Generation, Vegetation and Life would cease, and the Planets and Comets would not remain in their Orbs ' (p. 375).

Modern thermodynamics gives a different verdict ; but Newton's system was eternal and causes to prevent its decay must be sought out and investigated. To the causes of gravitation and of fermentation as such active principles, Newton presently added that of cohesion. And now he reached his well-known conclusion as to the origin and structure of material bodies :

> ' All these things being consider'd, it seems probable to me, that God in the Beginning form'd Matter in solid, massy, hard, impenetrable, moveable Particles, of such Sizes and Figures, and with such other Properties, and in such Proportion to Space, as most conduced to the End for which he form'd them ; and that these primitive Particles being Solids, are incomparably harder than any porous Bodies compounded of them ; even so very hard, as never to wear or break in pieces ; no ordinary Power being able to divide what God himself made one in the first Creation. While the Particles continue entire, they may compose Bodies of one and the same Nature and Texture in all Ages : But should they wear away, or break in pieces, the Nature of Things depending on them, would be changed. Water and Earth, composed of old worn Particles and Fragments of Particles, would not be of the same Nature and Texture now, with Water and Earth composed of entire Particles in the Beginning. And therefore, that Nature may be lasting, the Changes of corporeal Things are to be placed only in the various Separations and new Associations and Motions of these permanent Particles ; compound Bodies being apt to break, not in the midst of solid Particles, but where those Particles are laid together, and only touch in a few Points ' (p. 376).

And beyond the *vis inertiæ* and the passive laws of motion resulting therefrom, there were certain active principles that moved the particles, such as the principle of gravity and the principle that caused fermentation and the cohesion of bodies. These active principles, Newton insisted, were not occult qualities arising from specific forms, but general laws of nature; they were manifest qualities, whose causes only were occult, and occult only in the sense that they were not yet discovered. To explain phenomena in terms of unknown occult undiscoverable qualities was to put a stop to natural philosophy and told us nothing. He therefore preferred to propose these principles of motion, since they were of very general extent and derived from phenomena; and he would leave their causes to be found out (26).

What we have here set out shows that Newton had elaborated a coherent chemical philosophy. Its fundamental basis was the ancient atomic theory in which the transmutability of matter was implied. Whether water was the origin of all things, as Newton thought at the time when he was writing the 'Principia', hardly mattered when he had arrived at a full acceptance of ancient atomism as revived in his own time, an acceptance that is evident in the paper on the nature of acids. But Newton had made his own characteristic contribution in arguing from phenomena (and he insisted that his conclusions were so derived) that, between the particles of bodies, there was an exceedingly strong force of attraction reaching only to a short distance from the particles and producing the phenomena of chemistry; and he had, as we have shown, derived this from detailed evidence extending over wide ranges of chemical facts, which we have already sufficiently quoted above not to need any further repetition here. Nature, he had concluded, was 'very conformable to her self and very simple, performing all the great Motions of the heavenly Bodies by the Attraction of Gravity which intercedes those Bodies, and almost all the small ones of their Particles by some other attractive and repelling Powers which intercede the Particles'.

Newton's acceptance of atomism was an event in the history of chemistry. His influence on the work of Hales was admitted by that remarkable experimenter himself; Black might even be said to have been his disciple; Priestley was greatly attracted and inspired by the 'Queries', which he described as 'bold' and 'excentric'; and Dalton's whole work was permeated with Newtonianism. While Newton's influence on chemistry is part of his influence on the development of all the sciences, it derives also from his own chemical studies and the chemical philosophy that is scattered among his various publications.

The alchemical MSS. and notes.

From this close study of Newton's chemical philosophy as derived from his published works, it would be difficult to conclude that he was 'a credulous alchemist'. Rather, he appears as a pioneer in the study of chemical attraction, as in fact he appeared to many of those who came immediately after him and who held his 'Queries' in such high regard during the eighteenth century. But Newton copied out many long passages and made very extensive extracts from a large number of alchemical writings; and these are so numerous as to have led some authors to infer, from this evidence alone, that Newton was an 'adept' in the 'art' and devoted to the study of the alchemical transmutation of metals (of course, of the base metals into gold!). Too much significance is put upon these extracts. Many of them were from rare books; and many alchemical books gave useful information on the properties of the metals and their compounds. Moreover, Newton copied out passages from books that were not alchemical. And what chemist, even now, would not copy out, if he could lay hands upon them, the verses of 'Pearce the Black Monk upon y^e Elixir'? We might also ask what man of science would care to be judged by what he had copied out of books. Until we know what Newton's own opinion was about the material that he copied out from alchemical works, we have no warrant, however extensive that material may be and however uniform its content, for concluding that it reflected his own views on chemistry, especially when the chemical thought set forth in his published writings shows an advance towards modern chemistry rather than a reversion to the already discredited pursuit of gold-making. There is far more to be said in favour of the view that, since Newton's real interest lay in the study of the forces involved in chemical change, his object was to arrive at some general law of chemical attraction corresponding to his law of gravitational attraction and that he hoped to discover some numerical relationship for the attractive force between small particles of matter similar to that which he had discovered for the attractive force between large masses of matter. That he was not successful is but an indication of the greater complexity of the problem; mere mass is no longer a factor, and the attraction varies with the chemical nature of the mass, while chemical mass in the sense of chemical equivalent is a measure that, in Newton's day, was yet to come. At this moment we are far from a satisfying solution of this difficult problem.

Much the same may be said of Newton's correspondence about Boyle's recipe as of his extracts from alchemical writers. Part of the recipe dealt with the incalescence of gold with mercury; the production of heat was a phenomenon in which Newton was greatly interested; as for the alchemical aspects of the affair, Newton wrote to Oldenburg: 'For in good earnest, I have no opinion of it' (27). Credulity was not one of his failings.

Finally, we may quote from Newton's close friend, Dr. William Stukeley, although we may be unwilling to share his opinion of the science of chemistry:

> 'As to chymistry in general, we may very well presume Sir Isaac, from his long and constant application to that pyrotechnical amusement, had made very important discoverys in this branch of philosophy, which had need enough of his masterly skill, to rescue it from superstition, from

> vanity and imposture, and from the fond inquiry of alchymy and transmutation. By this means Sir Isaac carryed his inquiry very far downwards into the ultimate component parts of matter, as well as upwards towards the boundless regions of space; he has taken in to our knowledg a large province of universal nature, and put us in the way of acquiring more, if we have proper qualifications for it ' (28).

Stukeley was much in Newton's confidence; his opinion is not one that can be ignored; and he clearly did not suppose that Newton was given to ' the fond [*i. e.*, foolish] inquiry of alchymy and transmutation '.

Addendum: an anomalous barometric experiment.

In the early years of the Royal Society Huygens had noted that, when the Torricellian experiment was carried out with air-free water in the receiver of an air-pump and the receiver was evacuated, the water did not subside in the tube (29). In 1663 he demonstrated this before the Society, who had doubted the correctness of his observations. Boyle repeated the experiment with air-free mercury and obtained columns of mercury of 34, 52, 55 and even 75 inches in height under vacuum in similar apparatus; and he found that the mercury would remain suspended at the top of the tube even when the tube was removed to the open air, but that the admission of the slightest bubble of air caused the mercury to fall to its usual height of 27 or 28 inches. This failure or anomalous result of the Torricellian experiment was very difficult to explain; theories were advanced by Huygens and Chérubin and an explanation was promised by Hooke, but Boyle and his colleagues failed to find any satisfactory solution. To-day, of course, we recognize that clean mercury ' wets ' clean glass, and Manley (30) has been able to demonstrate the concavity of the mercury-glass meniscus. This latter phenomenon may, in fact, be observed in high-vacuum apparatus in such contrasted conditions as provide a convex mercury-glass meniscus in an adjoining position, *e. g.*, in the low-pressure limb of a McLeod gauge or a gas-burette, where in the high-pressure limb of the gauge or the burette the meniscus is convex, as the present writer has frequently noted; and it may be readily and rapidly demonstrated by subjecting to high vacuum a narrow glass tube containing clean mercury, the tube being tapped or shaken to facilitate the removal of the air below the mercury. It is not without interest to note here that in the ' Opticks ' Newton had explained the anomalous Torricellian experiment as due to cohesion between glass and air-free mercury and to remember that by cohesion he meant a short-ranged force of attraction between the particles of bodies (p. 365).

References.

(1) William Stukeley, ' Memoirs of Sir Isaac Newton's Life ', edited by A. Hastings White, London, 1936, pp. 35–6. ' Ropesly ' (now Ropsley) is probably meant rather than ' Roperly '; and ' Bassingthorpe ', near

Grantham, is more probable than 'Basingham' (now Bassingham), near Lincoln. 'Holland' is, of course, the Lincolnshire district of that name.

(2) Andrade has shown ('Nature', 1935, cxxxv, p. 360) that the 'collection' of receipts that D. E. Smith supposed to have been made by Newton ('Isaac Newton 1642–1727', edited by W. J. Greenstreet, London, 1927, pp. 16 *ff.*) was no collection of Newton's own, but almost entirely an extract from the third edition (1654) of William Bate's 'The Mysteries of Nature and Art', and that the date of the first entries was 1659 rather than 1655–8.

(3) Brewster's 'Memoirs of . . . . Sir Isaac Newton', Edin., 1855, vol. i, p. 32, foot-note 1. 'Oyle pink' is not easy to identify, but we may hazard a guess that it may have been an extract of the flower of that name for use as an early 'indicator', such 'syrups' being then in use, *e. g.*, by Boyle. The actual copy of the 'Theatrum Chemicum' owned by Newton, the edition of 1659 in six volumes, was recently offered for sale by a London bookseller.

(4) *Ibid.*, pp. 388–9.

(5) L. T. More, 'Isaac Newton', New York and London, 1934, p. 52.

(6) See Partington, 'Annals of Science', 1936, i, p. 368.

(7) 'Isaaci Newtoni Opera quæ exstant omnia', edited by Samuel Horsley, London, 5 vols., 1779–85: see vol. iv, pp. 273–4, where Newton describes some of his results in letters written in January of 1671/2. Brewster (*op. cit.*, vol. ii, p. 361, foot-note 1) reports also records of similar experiments on Dec. 10–19, 1678, and of others in 1679 and 1680. We might also recall here that there is an alloy that still bears Newton's name, 'Newton's alloy' or 'Newton's fusible metal', which consists of 8 parts of bismuth, 5 parts of lead and 3 parts of tin, and which melts at the B.P. of water. Readers who are unfamiliar with the use of such dates as 1671/2 might like to be reminded that this is due to the moving of New Year's Day from March 25 to January 1, whereby the intervening period from January 1 to March 24 was, for example, part of the year 1671, but is now to be reckoned as 1672: this reform was legalized for the year 1752 by an Act of Parliament passed in 1751, which also removed the 'eleven days' from the month of September, 1752, in order to allow for accumulated errors in the calendar. The letters O.S. (Old Style) are frequently used to indicate that a date has suffered neither of these corrections.

(8) Birch, 'History of the Royal Society of London' etc., London, 1756–7: see vol. iii, pp. 249–50.

(9) Anaximandros (611–546 B.C.) sought the origin of all things in an indefinite ætherial principle (ἄπειρον).

(10) Birch, *op. cit.*, vol. iii, p. 251.

(11) Brewster, *op. cit.*, vol. i, p. 409.

(12) *Ibid.*, vol. ii, p. 93.

(13) *Ibid.*, pp. 95–6.

(14) First published at Basle in 1556. Newton possessed a copy of the third edition of 1621. See R. de Villamil, 'Newton: The Man', London, n. d. (probably 1931), p. 63. The catalogue of Newton's Library, as given by Col. de Villamil, shows that Newton had acquired a very representative collection of chemical works rather than an alchemist's library. It is curious that this author should have drawn attention to the absence of the English classics, including the works of Milton (p. 9), as 'Milton's Works (Poetical) 2 Vol., 4to. 1720' is included on p. 86.

(15) Brewster, *op. cit.*, vol. ii, pp. 361–2.

(16) The text of this is given by Brewster, *op. cit.*, vol. ii, pp. 535–6.

(17) Professor J. R. Partington, M.B.E., has kindly informed me that Newton's copy of van Helmont's 'Opera' (1667), which is in his possession, is identical in content with the 'Ortus Medicinæ' (1648), the work of van Helmont usually referred to.

(18) 'Sir Isaac Newton's Mathematical Principles of Natural Philosophy . . . . Translated into English by Andrew Motte in 1729 . . . . revised . . . . by Florian Cajori', etc., Cambridge, 1934, pp. 529–30.

(19) *Ibid.*, p. 542.

(20) 'Principia', 1687, Lib. III, p. 506. See also McKie, 'Ambix', 1938, i, 162–3.

(21) The author's thanks are due to Professor E. N. da C. Andrade, F.R.S., for his kindness in lending a copy of this rare supplementary volume in his possession.

(22) More, *op. cit.*, p. 163.

(23) *Ibid.*, pp. 163–4.

(24) See Partington and McKie, 'Annals of Science', 1938, iii, pp. 337–71.

(25) Lemery's theory will be found in the English version of his work by James Keill, 'A Course of Chymistry' etc., 3rd edition, London, 1698, pp. 165–6. It occurs also in earlier works by Lemery, but we quote this edition, as it was probably the source used by Newton, since he possessed a copy of it (de Villamil, *op. cit.*, p. 83).

(26) It might be noted here that Newton had already given a hint of the conclusion he now presented. In the General Scholium added to Book III of the second edition (1713) of the 'Principia', he had said: 'And now we might add something concerning a certain most subtle spirit which pervades and lies hid in all gross bodies; by the force and action of which spirit the particles of bodies attract one another at near distances, and cohere, if contiguous . . . .'. But these were things that could not be explained in a few words, and there was not enough experimental evidence available for accurate determination and demonstration. Moreover, Newton had just delivered himself of his famous cautionary dictum on the use of hypotheses in science: 'But hitherto I have not been able to discover the cause of those properties of gravity from phenomena, and I frame no hypotheses; for whatever is not deduced from the phenomena is to be called an hypothesis; and hypotheses, whether metaphysical or physical, whether of occult qualities or mechanical, have no place in experimental philosophy. In this philosophy particular propositions are inferred from the phenomena, and afterwards rendered general by induction'. See Cajori's edition of Motte's translation of the 'Principia', cited above, p. 547.

(27) See Boyle's 'Works', edited by T. Birch, London, 1772, vol. vi, p. 286; Brewster, *op. cit.*, vol. ii, pp. 120–22, 376 (foot-note 1), and 461–2; and More, *op. cit.*, pp. 162 and 371–4.

(28) Stukeley, *op. cit.*, p. 60.

(29) See McKie, 'Chérubin d'Orléans: A Critic of Boyle', in 'Science Progress', 1936, xxxi, p. 55. A further study of the history of this problem is in course of preparation.

(30) Manley, 'Philosophical Magazine', 1928, v, p. 958.

LXXXIX. *Newton and the ‘Principia.’*

By Allan Ferguson, M.A., D.Sc.*

[Received November 26, 1942.]

The year 1942 is a year of centenary celebrations of various orders. 1642 saw the death of Galileo and the birth of Newton, 1742 saw the death of Halley. In the story of the development of physical science in the seventeenth century these names are indissolubly linked. Galileo contributed to the laying of the foundations of the imposing structure of what is now termed classical dynamics, Newton raised the structure, and endowed it with a stability which seemed, in the eyes of the eighteenth and nineteenth century men of science, to make it capable of enduring for all time ; and Halley, who has first-rate achievements to his credit in more than one branch of physical science, will ever be remembered as the unselfish friend, to whose learning, tact, editorial and monetary aid we owe it that the ‘ Principia ’ was ever written and published. De Morgan, who does not scatter praises with a lavish hand, well says, “ The part which Halley had in the matter would alone immortalize his name. He found out the ability of Newton to write such a work, prevailed upon him to write it, took charge of the publication, prevented the author from materially mutilating it in disgust, paid the expenses of printing, at a time when, owing to his father’s death and consequent litigation, he had nothing to spare (which never happened to him before or after), gave a copious explanation of it in the ‘ Philosophical Transactions,’ and is generally admitted to have been for a long time the only person in Europe who showed that he thoroughly appreciated the value of the work, and knew the place it must occupy in the history of discovery.”

But this is to anticipate the romantic story of the writing and publication of the modest quarto volume, whose title page reads : *Philosophiæ Naturalis Principia Mathematica. Auctore Is. Newton, Trin. Coll. Cantab. Soc., Matheseos Professore Lucasiano, et Societatis Regalis Sodali. Imprimatur. S. Pepys, Reg. Soc. Præses, Julii* 5, 1686. *Londini, Jussu Societatis Regiæ ac Typis Josephi Streater. Prostat apud plures Bibliopolas. Anno MDCLXXXVII.* Let us trace, as quickly as may be, the sequence of events which led up to the publication of a work which has few, if any, rivals in the history of the development of scientific thought.

The hamlet of Woolsthorpe lies off the Great North Road, a few hundred yards north-west of Colsterworth, and is about six miles south of the market town of Grantham in Lincolnshire. The little river Witham, which rises near Woolsthorpe, flows northward between the two villages, and at the small manor farm of Woolsthorpe on Christmas

---

* Communicated by the Author.

Day 1642 (O.S.) was born Isaac, the posthumous son of Isaac Newton, who had owned and farmed the manor, an estate of the yearly value of some £30. In 1645 Newton's mother became the wife of the Reverend Barnabas Smith, rector of North Witham, and the care of Newton devolved upon the child's maternal grandmother. After a period of day schooling at Skillington and Stoke, hamlets about a mile north of Woolsthorpe, he was sent at the age of twelve to the school at Grantham, where he boarded at the house of one Mr. Clark, an apothecary. In 1656, Newton's mother, again widowed, returned to Woolsthorpe with the three children of her second marriage, and Newton, recalled from Grantham, entered on a career as farmer and grazier. Fortunately, his relatives soon saw that Newton's genius did not lie in the direction of tilling the land and marketing cattle, and he returned to school at Grantham, where he was to be prepared for Cambridge, on the recommendation of a maternal uncle, the Reverend W. Ayscough. Such records of his youth as have come down to us show that the young Newton was a lad of parts, though there is little to indicate the tremendous sweep of his mature mind. What does come out clearly is that he was a skilful workman with his hands, possessing considerable aptitude for the construction of kites, model windmills, water-clocks and sundials, the accuracy of the clocks and dials being, if local testimony has any value, of a high order. In this respect, certainly, the child was father to the man; and it is even to-day hardly recognized widely enough that Newton was pre-eminently an experimenter, and the common picture of the absent-minded mathematician needs considerable qualification.

Newton was admitted to Trinity College as a subsizar in June 1661, and matriculated on July 8 of that year. The accounts of his undergraduate studies are scanty, but it is on record that, having no knowledge of Euclid, he acquired a copy for the purpose of elucidating some of the propositions in a book on astrology bought at Stourbridge fair. He found the Euclidean problems and theorems so self-evident that he considered Euclid a trifling book, and paid no further attention to it. He was admitted scholar of Trinity in April 1664, and Barrow, who examined him in Euclid, found him weak in that subject—a curious comment on one who, in later years, was to handle geometrical methods with unexampled power. Newton graduated bachelor in January 1665, but the *ordo senioritatis* of the candidates for that year is missing, and we have, therefore, no record of the relative merit of Newton's performance in the examination. Stukeley puts on record as a tradition holding while he was a student at Cambridge that "when Sir Isaac stood for Bachelor of Arts degree, he was put to second posing, or lost his groats, as they term it; which is looked upon as disgraceful." *

---

* The editor of Stukeley's 'Memoirs of Sir Isaac Newton's Life' (p. 53) remarks concerning the strange phrase "lost his groats"—"'groats' not 'greats.' This curious expression seems quite lost, unrecorded, and unexplained." Perhaps some reader of the 'Philosophical Magazine' can throw light on its origin.

The plague visited Cambridge and much of the years 1665 and 1666 were spent by Newton at Woolsthorpe, where he meditated many matters, including the question of gravitation. It must not be assumed that Newton was alone in pondering on this problem ; it was in the minds of many men, and as De Morgan has, not inaptly, remarked, " the notion of gravitation was *not new* ; but Newton *went on*." The story of the falling apple is a pleasing tale, and rests on fair authority. Conduitt, husband of Newton's niece, and Mrs. Conduitt herself are primary authorities ; R. Greene (*The Principles of the Philosophy of the Expansions and Contractive Forces* . . . . cited by De Morgan) had the story from Martin Folkes ; Isaac D'Israeli, probably on his own initiative, makes the apple strike him a smart blow on the head ; Pemberton merely remarks, " As he sat alone in a garden, he fell into a speculation on the power of gravity : that as this power is not found sensibly diminished at the remotest distance . . . . to which we can rise . . . . it appeared to him reasonable to conclude that this power must extend much farther than was usually thought ; why not as high as the moon, said he to himself ? " Stukeley, who is a primary authority, is precise. He says (p. 19) : " On 15 April 1726 I paid a visit to Sir Isaac at his lodgings . . . . dined with him and spent the whole day with him, alone. . . . . After dinner, the weather being warm, we went into the garden and drank thea, under the shade of some appletrees, only he and myself. Amidst other discourse, he told me, he was just in the same situation, as when formerly, the notion of gravitation came into his mind. It was occasion'd by the fall of an apple, as he sat in a contemplative mood. Why should that apple always descend perpendicularly to the ground, thought he to him self. Why should it not go sideways or upwards, but constantly to the earth's centre ? Assuredly, the reason is that the earth draws it. There must be a drawing power in matter : and the sum of the drawing power in the matter of the earth must be in the earths center, not in any side of the earth. Therefore dos this apple fall perpendicularly, or towards the center. If matter thus draws matter, it must be in proportion of its quantity. Therefore the apple draws the earth, as well as the earth draws the apple. That there is a power, like we here call gravity, which extends itself thro' the universe."

Stukeley's account must be allowed some weight ; his MS. of date 1752, was written from records made in Newton's lifetime—as he says in his preface, " I . . . . have not been overhasty * in printing, what was wrote 27 years ago." The apple story may, then, be regarded as having a sufficient basis in fact. What is more important is the reason behind the twenty years' delay in the publication of Newton's first calculation of the moon's acceleration towards the earth based on the inverse square law and its comparison with the actual value. The explanation which was commonly accepted until half a century ago and is, indeed, still current, may be told in Pemberton's own words (Preface to his *View of Sir Isaac Newton's Philosophy* '): "In this computation, being absent from books, he took the common estimate in use among geographers and our seamen, before *Norwood* had measured the earth, that 60 English miles were contained in one degree of latitude on the surface of the earth. But as this is a very faulty supposition, each degree containing about 69½ of our miles, his computation did not answer expectation; whence he concluded, that some other cause must at least join with the action of the power of gravity on the moon. On this account he laid aside for that time any further thoughts upon this matter. But some years after, a letter which he received from Dr. *Hook* . . . . gave occasion to his resuming his former thoughts concerning the moon; and *Picart* in *France* having lately measured the earth, by using his measures the moon appeared to be kept in her orbit purely by the power of gravity; and consequently, that this power decreases as you recede from the center of the earth in the manner our author had formerly conjectured."

---

* Indeed, he was not over-hasty. He died in 1765, leaving an immense mass of papers, which includes two forms of the 'Memoirs' each bearing date 1752. The Memoirs appeared in print only a little more than six years ago ('Memoirs of Sir Isaac Newton's Life,' by William Stukeley, M.D., F.R.S. 1752. Edited by A. Hastings White. London, Taylor & Francis, 1936).

Thus far Pemberton, and his exposition is practically that followed by most historians until the later years of the nineteenth century, when it was pointed out that Newton at this stage of his work had not arrived at the theorem on the attraction of a solid sphere on an external particle which showed that the Earth attracted the apple as if the Earth's mass were concentrated at its centre, and the view was put forward that it was for this reason, rather than for that given by Pemberton, that Newton laid aside his papers on gravitation *.

Writers of modern date are divided in their allegiance to these different views. Sir Robert Ball (*Great Astronomers*) does not discuss the point. After describing Newton's meditations at Woolsthorpe, he proceeds: "Newton was able to calculate . . . . what the attractive power of the earth must be, so that the moon shall move precisely as we find it to move. It then appeared that the very power which makes an apple fall at the earth's surface is the power which guides the moon in its orbit."

---

* Newton's calculations may be summarized thus in modern language. If the radius of the moon's orbit be R, and her period T, her earthward acceleration will be $f=v^2/\mathrm{R}=4\pi^2\mathrm{R}^2/\mathrm{T}^2\mathrm{R}=4\pi^2\mathrm{R}/\mathrm{T}^2$. Put $\mathrm{R}=60a$, where a is the radius of the earth. This was the current estimate in Newton's day; the modern estimate is $60{\cdot}27a$. Put $a=3960$ miles and $\mathrm{T}=39{,}343$ minutes and we find $f=0{\cdot}00893$ ft. sec.$^{-2}$. According to the inverse square law $f/g=(\mathrm{R}/60{\cdot}27\mathrm{R})^2$, with g taken as 32·2 ft. sec.$^{-2}$ gives $f=0{\cdot}00887$ ft. sec.$^{-2}$. The agreement is good; but if we use, as it is suggested Newton used, a value of the earth's radius based on an estimate of 60 miles to a degree, the word "agreement" is completely out of place.

If we assume circular motion, the inverse square law follows from a consideration of Kepler's third law.

Berry (*Short History of Astronomy*) remarks : " It is probable that he used a seriously inaccurate value of the size of the earth, having overlooked the measurements of Snell and Norwood ; it is known that even at a later stage he was unable to deal satisfactorily with the difficulty . . . . as to whether the earth might for the purposes of the problem be identified with its centre ; and he was of course aware that the moon's path differed considerably from a circle."

Brodetsky (*Sir Isaac Newton*) adopts Pemberton's view, that Newton calculated the moon's distance as $60a$, with the radius of the earth determined on the basis of 60 statute miles to a degree. He makes Newton's calculation and finds a discrepancy between observed and calculated values of nearly 16 per cent. " Not a word did Newton say of all this to a living soul . . . . about sixteen years later he discovered . . . . the cause of his failure—the simple fact that the value of R he had assumed was too small, and that he should have taken it by about fifteen per cent. greater."

Buckley (*Short History of Physics*) quotes Newton's remark concerning the agreement of the calculated and observed results answering " pretty nearly," points out that the discrepancy is of the 15 per cent. order, and that the discrepancy disappears when Picard's value of the length of a degree is used. " The discrepancy between the two results now disappears and he regarded his ideas as fully confirmed."

Cox (*Mechanics*) makes the calculation on the sixty miles to a degree basis, shows the usual large discrepancy and remarks that " Newton's behaviour in face of this disappointing result is as marvellous an instance of scientific reserve, as his daring guess was of scientific imagination." He states that Newton's laid aside the subject until he heard of Picard's measurements, when he recalculated his results and " it is said that as it became clear that the result would accord with theory, his excitement was so great that he could not see the paper to finish his work " *.

Sir Wm. Dampier (*History of Science*) after criticizing the " Pemberton " view, goes on to say, " A more likely reason for Newton's delay in publishing his calculations was pointed out in 1887 by J. C. Adams and J. W. L. Glaisher," and discusses the difficulty involved in calculating the attraction of the Earth on a small body near its surface. " This was probably the chief reason why Newton put aside his work in 1666."

Libby (*Introduction to the History of Science*) gives both views due measure. " Two doubts had made his meticulous mind pause—one, of the accuracy of the data in reference to the measurement of the meridian, another, of the attraction of a spherical shell upon an external point."

Macpherson (*Makers of Astronomy*) mentions the difficulty involved in the calculation of the attraction of a sphere, but assumes that Newton's chief difficulty was due to his use of a wrong value for the radius of the

---

* It is only fair to say that Cox has a footnote, " There seems to be no authority for this particular story, (of the "excitement"), first given by Robison in 1804."

Earth. He quotes the 15 per cent. error and says, "the discrepancy was so great that Newton appears to have concluded that he was on the wrong track. So he laid aside the investigation and turned to there branches of science."

Pledge (*Science since* 1500) points out the three types of case with which a theorist must deal—two points and elliptic orbit (sun and planet), two finite spheres (moon and earth), and earth and small body near its surface, but does not discuss the points under review.

Singer (*Short History of Science*) remarks that, "when Newton first approached this problem (1666) he found that the moon's fall was but seven-eighths of what he expected. . . . . He had years of labour before he could show that the attraction of a spherical body on an external point was as if the spherical body were concentrated at its centre (1685)." Later he notes, of Picard's measurement, "it was at first missed by Newton, who, in ignorance of it, abandoned for some years his calculations, based on earlier measurements, seeking to identify gravity as the force that kept the moon and planets in their orbits."

H. M. Taylor (Encyc. Brit. 11th ed., art. *Newton*) takes the 'Pemberton' view.

R. de Villamil (*Newton : the Man*) has some shrewd incidental remarks. He expresses his disbelief in Newton's ignorance of the fairly accurate measurements of a degree which had been made earlier in the century, notes that in Newton's words the calculations were found to answer "pretty nearly," and after asking the question "Why did not Newton publish this result ?" gives the simple answer that Newton hated self-advertisement, and that he did not at that time *know* "what was the true 'distance' to be taken."

These may be taken as a fair sample of views lately current, and they are to be divided into two main classes; those which adopt the explanation given by Pemberton, and those which assume Newton to be held back rather by his doubts concerning the true measure of the distance between the apple and the earth's centre of attraction. It is to be remembered that we have nowhere a note on Newton's calculations —otherwise the argument would never have arisen. His own words are: "And the same year (1666) I began to think of gravity extending to y^e orb of the Moon and . . . . I deduced that the forces w^{ch} keep the Planets in their Orbs must (be) reciprocally as the squares of their distances from the centres about w^{ch} they revolve, and thereby compared the force requisite to keep the Moon in her Orb with the force of gravity at the surface of the earth, and found them answer pretty nearly." (Portsmouth draft memorandum, quoted by W. W. R. Ball, *Essay*).

Again (Newton to Halley, letter, June 20, 1686. Brewster (i), 439): ". . . . in one of my papers writ . . . . above fifteen years ago . . . . the proportion of our gravity to the moon's conatus recedendi a centro terræ is calculated, though not accurately enough."

"*Answer pretty nearly*"; "*though not accurately enough.*" These phrases would hardly be used by one of Newton's cautious and cool temperament of a calculation labouring under a fifteen or sixteen per cent. error. And that error would not be improved had Newton used another common measure—the mile of a thousand paces, with five (English) feet to the pace. What estimate did Newton use ? * What literature was easily accessible, and likely to be read by him at Cambridge ? Hues, *Tractatus de Globis* (1594) has a chapter " *Of the Circumference of the Earth, or of a Greater Circle*; *and of the Measure of a Degree*," in which he discusses the measurements of authorities, ancient and modern, and gives a table of the " circuit of the whole earth," in which the extreme figures are 172,000 and 252,000 " furlongs." (*Furlong* and *stadium* are convertible in Hues' diction.)

Wright (*Certaine Errors in Navigation*, 1610 and later), describes a determination of the earth's radius in which the possibilities of large errors are abundantly obvious, and which gives almost exactly 60 statute miles to the degree.

In 1617 Snell (see Encyc. Brit. 11th edn. *Earth, figure of the*) measured a base line near Leyden, and carried out therefrom a series of triangulations. His estimate of the length of a degree is given in toises (toise =6·4 Eng. feet approx.), and in English measure one degree=66·9 statute miles.

But by far the most accurate of the pre-Picard determinations is that made by Norwood (*The Seaman's Practice* . . . . 1636. A very popular work, which passed through several editions before 1666). I quote the description of Norwood's work given by Bonnycastle ('Astronomy,' 6th edn., 1811). It will serve to illustrate the interest taken in geodesy, and to show that nineteenth century commentators on Newton had access, in popular treatises, to fairly full accounts of Norwood's measurements—I do not discuss the detailed accuracy of the description. " But the person who engaged in this enterprise with the most success, was our countryman, Mr. Richard Norwood. In the year 1635 he took the sun's altitude, when it was in the summer solstice, both at London and York, with a sextant of five feet radius, and by that means found the difference of latitude between these two cities to be two degrees and twenty-eight minutes. He then measured their distance in as exact a manner as he was able ; and having taken into account all the turnings and windings in the road, with the ascents and descents, he reduced it to an arc of the meridian and found it to contain 12,849 chains ; and this . . . . gave him 5209 chains to a degree." Norwood's value is, in fact, almost exactly 69·5 statute miles to the degree.

Newton had, in effect, a choice between 60 miles to the degree, and 67 or 69 miles. Is there any doubt that, in view of his own description of his early calculation, he used the more exact figure ? And that he laid his calculations aside, either because he saw the serious nature of the assumptions involved in treating the earth as a particle, or, what is very probable, because he became deeply involved in his optical researches ?

---

* The question is discussed briefly by W. W. R. Ball (*Essay*, p. 15) and with a great wealth of detail by Cajori (article in *Sir Isaac Newton, 1727–1927. A Bi-centenary Evaluation*. . . . p. 127).

The story of the writing and publication of the *Principia* must be compressed into as small a space as may be *. As we have seen, it is not a very difficult matter to deduce the inverse square law from a consideration of circular motion and Kepler's third law. Hooke, Halley and Wren had each apparently accomplished this feat. But the inverse problem—given the law of force, what will be the path of a planet ?—proved too much for all of them. In a well-known letter from Halley to Newton we find the statement that "in January 1683–4 I, having from the consideration of the sesquialterate proportion of Kepler, concluded that the centripetal force decreased in the proportions of the squares of the distances reciprocally, came on Wednesday to town (from Islington), where I met with Sir Christopher Wren and Mr. Hooke and falling in discourse about it, Mr. Hooke affirmed that upon that principle all the laws of the celestial motions were to be demonstrated, and that he himself had done it." In the same letter he chronicles an offer made by Wren in a sporting spirit which we do not usually associate with his name. We read that "Sir Christopher, to encourage the enquiry said that he would give Mr. Hooke some two months' time to bring him a convincing demonstration thereof, and besides the honour, he of us that did it should have from him the present of a book of forty shillings. Mr. Hooke then said he had it, but that he would conceal it for some time, that others trying and failing might know how to value it when he should make it public. However, I remember that Sir Christopher was little satisfied that he could do it, and though Mr. Hooke then promised to show it him, I do not find that in that particular he has been so good as his word."

It was Halley who took the problem to Newton, to find that Newton had long ago solved this and other problems of planetary motion, and it was Halley's enthusiasm that overcame Newton's constitutional aversion to publication, and drove him to put forward his views in connected form. On April 28, 1686, Dr. Vincent laid before the Royal Society the MS. of the first book of the 'Principia,' and on May 19 the Society resolved that "Mr. Newton's *Philosophiæ Naturalis Principia Mathematica* be printed forthwith in quarto in a fair letter." On June 2 we hear a different note ; the Council orders that "Mr. Newton's book be printed, and that Mr. Halley undertake the business of looking after it and printing it at his own charge" ; and Halley undertakes the task.

The Society was, in fact, in financial straits. It had, some time before, published "Willoughby de Historia Piscium," doubtless a learned volume, and one which then excited as much attention as a similar volume would do to-day. The Society made desperate efforts to minimise its loss, and the phrase "books of fishes" recurs in the Council minutes of the period like “ the owercome o’ a sang.” The salaries of their officers fell into arrears, and the Council resolved to pay the unlucky philosophers in “ books of fishes.” Halley was agreeable ; Hooke more cautiously desired six months time to consider the acceptance of such payment ! Again, when the Society asked Halley to undertake the measurement of a degree of longitude, it was resolved that he “ be given £50, or fifty books of fishes.”

---

* The account which follows is taken from an article on *Edmond Halley* by the present writer ('Nature,' cxxxi. p. 153 (1933)).

Newton, who could on occasion show what in a lesser man would be called peevishness, did not make Halley’s task any easier by desiring to suppress the MS. of the third book of the ‘ Principia,’ on hearing that Hooke was putting forward claims to priority. But Halley’s perseverance triumphed over all difficulties and on July 5, 1686, the book received the *Imprimatur* of the then president, Samuel Pepys ; it was published about the middle of 1687.

What are the pattern and scope of the ‘ Principia ’ ? The question is not rhetorical. The mid-nineteenth century candidate for mathematical honours in the University of Cambridge was expected to read a few sections of the first book ; W. W. R. Ball, writing in 1893, could describe the book as “ easily accessible.” But a knowledge of the ‘ Principia ’ is no longer demanded of candidates sitting for the Mathematical Tripos, and it is doubtful whether the phrase “ easily accessible ” applied in the eighteen-nineties, if we interpret the phrase as involving not merely physical access to the book itself, but mental access implied in a knowledge of the Latin tongue and in some measure of skill in Newton’s geometrical methods. It certainly does not apply now, even if we pay due regard to the fact that Professor Cajori has recently re-published and very fully annotated Motte’s English translation, which was first published in 1729. And in the matter of physical access only, there are many libraries specializing in books on physical science (some in Newton’s own University town) which do not possess a copy of the ‘ Principia,’ whether in Latin or in English translation.

Several synopses of the ‘ Principia ’ have from time to time been made. The faithful Halley gave a full review in the current *Philosophical Transactions*, in which he expounded the principal themes and methods employed in the book. De Morgan (*Penny Cyclopædia*, article *Principia*) and W. W. R. Ball (*Essay on Newton’s Principia*) have given very detailed and complete analyses of the three books of the ‘ Principia.’ Ball (*Cambridge Papers*) has also made a shorter synopsis of the work.

The three editions of the ‘ Principia ’ published in Newton’s lifetime vary considerably in matters of detail, but they may be briefly described as made up of the dedication to the Royal Society, prefaces, and verses by Halley in Newton’s honour. Then follow eight definitions and three axioms or laws of motion. The first book, on the motion of bodies in unresisting media, is divided into 14 sections containing some 98 propositions, besides various lemmas and scholia. The second book, on motion in resisting media, is divided into nine sections, containing 53 propositions with lemmas and scholia. The third book on the system of the World comprises a preface, four rules of reasoning in philosophy and 42 propositions. The book concludes with a *scholium generale*, which is so eloquent of Newton in his exalted moments, that no apology is needed for presenting it *in extenso*; before doing so, we consider, in a little more detail, the topics discussed in the three books of the 'Principia.'

DEFINITIONS.

Eight in number, including mass (defined as volume × density), momentum, vis inertiæ, vis impressa (impressed force) centripetal force, the absolute magnitude of a centripetal force, accelerating force, moving force.

Scholium on time, space, and motion.

AXIOMS (LAWS OF MOTION).

The three laws of motion and six corollaries dealing with the composition and resolution of forces, the conservation of momentum and of the motion of the C.G. of a system, the effect on the relative motions of bodies (*a*) of motion of the space in which they exist, (*b*) of application of equal accelerations to each of them.

A scholium is added dealing with the laws of motion, which contains, *inter alia*, some account of experiments on impact.

BOOK I.—ON UNRESISTED MOTION.

Section I.—On Prime and Ultimate Ratios.

Consists of 11 lemmas and a scholium, dealing with Newton's method of mathematical attack, and a comparison of the methods of limits, of indivisibles and of exhaustions.

Section II.—On Centrifugal Forces.

Propositions 1–10, lemma 12, and a scholium.

Contains the propositions that if the force is centripetal, equal areas are described in equal times by the radius vector; and conversely; the calculation of centripetal force for uniform circular motion; given velocities at three points of the orbit to find the position of the centre of force; the law of force when a body describes an equiangular spiral about the pole; or an ellipse about the centre; and other propositions.

Section III.—Motion in Conics about the Focus.

Propositions 11–17, lemmas 13–14, and a scholium.

Contains the propositions on the law of force for an elliptic orbit; a hyperbolic orbit; a parabolic orbit—in each case described about the focus; in elliptic orbits the periodic times are in the sesquiplicate ratio of the major axes; given initial position and velocity under centripetal force varying according to the inverse square law, to determine the orbit; and other propositions.

Section IV.—On the Determination of Conic Sections, given the Focus.

Propositions 18–21, lemmas 15–16, and a scholium.

Section V.—On the Determination of Conic Sections, no Focus being given.

Propositions 22–29, lemmas 17–27, and scholia.

Section VI.—On the Determination of the Motion, the Orbit being a given Conic.

Propositions 30–31, lemma 28, and a scholium.

Contains the propositions, given a body moving in a parabolic orbit, the force being directed to the focus, to find its position at any given time; and a similar proposition for an elliptic orbit.

Section VII.—On Rectilinear Ascent and Descent of Bodies.

Propositions 32–39.

Section VIII.—On the Determination of an Orbit for a Body moving under any Law of Centripetal Force.

Propositions 40–42.

Section IX.—On the Motion of Bodies in Moveable Orbits and on the Motion of the Apsides.

Propositions 43–45.

Section X.—On the Motion of Bodies in given Surfaces and on Pendular Motion.

Propositions 46–56 and a scholium.

Contains the propositions, given a plane, and an external centre of force, the law of force being any whatsoever, to find the motion parallel to the plane; given a similar system and the law of force being as the distance, the orbit parallel to the plane will be an ellipse; to constrain a body to oscillate in a given hypocycloid; (given that the quadrature of curvilinear figures is possible) to find the law of force under which bodies oscillating in given curves will perform isochronous oscillations; and other propositions.

Section XI.—On the Motion of Bodies mutually attracted.

Propositions 57–69 and a scholium.

Contains the propositions that two mutually attracting bodies describe similar figures about their common centre of gravity and about each other; on the relation of the periodic times about the common centre of gravity, and of one body about the other, considered fixed; (what is commonly known as) the problem of the three bodies; and other propositions.

Section XII.—On the Attractions of Spherical Bodies.

Propositions 70–84, lemma 29, and scholia.

Contains the propositions that a particle placed anywhere inside a uniform gravitating spherical shell is in equilibrium; a spherical shell attracts a particle placed at an external point as if its mass were concentrated at its centre; and other propositions including the well-known theorems on the attractions of solid homogeneous spheres, or spheres whose density is a function of the distance from the centre.

Section XIII.—On the Attractions of Non-spherical Bodies.

Propositions 85–93, and scholium.

Contains propositions on the attraction of a uniform circular lamina at any point on its axis; the attraction of a solid of revolution at any point on its axis (with corollaries giving applications to a cylinder and a spheroid); and other propositions.

Section XIV.—On the Motion of Particles acted on by Centripetal Forces tending to the various Parts of a Large Body.

Propositions 94–98 and scholia.

Contains propositions on the sine law of refraction; on the ratio of the velocities of the particle before and after refraction; on the condition for total reflexion; on the construction of a surface of separation between two media such that all particles issuing from one point shall be refracted to another; and to construct a lens which shall have this property.

BOOK II.—ON RESISTED MOTION.

Section I.—On the Motion of Bodies when the Resistance of the Medium is proportional to the Velocity.

Propositions 1–4, lemma 1, and scholium.

Section II.—On Motion when the Resistance is Proportional to the Square of the Velocity.

Propositions 5–10, lemma 2 (on fluxions), and a scholium which varied from edition to edition, and bears on the Newton-Leibniz controversy.

Section III.—On Motion when the Resistance is made up of Two Terms —One Varying on the Velocity, the Other Varying as the Velocity Squared.

Propositions 11–14 and scholia.

Section IV.—On Spiral Motion in a Resisting Medium.

Propositions 15–18 and lemma 3.

Section V.—On the Density and Compression of Fluids and on Hydrostatics.

Definition of a fluid, propositions 19–23, and scholia.

Section VI.—On the Resisted Motion of Pendulums.

Propositions 24–31, and long scholium generale, dealing in detail with Newton's experimental results.

Section VII.—On the Motion of Fluids and Resistance of Projectiles.

Propositions 32–40, lemmas 4–7, and scholia. Scholium appended to prop. 40 gives details of 14 experiments on fluid resistance.

Section VIII.—On (Undulatory) Motion Propagated through a Fluid.

Propositions 41–50 and a scholium which gives the "Newtonian" value for the velocity of sound in air as 979 feet per second.

Section XIV.—On the Circular Motion of Fluids.

Hypothesis on the resistance of a viscous fluid, propositions 51–53,

BOOK III.—ON THE SYSTEM OF THE WORLD.

Introduction, Regulæ Philosophandi, Phenomena, Propositions 1–42, Corollaries and Scholia, Hypotheses 1_2, and Scholium Generale.

The "rules of reasoning in philosophy" are four in number and, representing, as they do, Newton's mature views, they are here given in full. The fourth does not appear in the second edition of the 'Principia'; it appears in the third (and subsequent) editions.

Regulæ Philosophandi.

1. We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances.
2. Therefore to the same natural effects we must, as far as possible, assign the same causes.
3. The qualities of bodies, which admit neither intensification nor remission of degrees, and which are found to belong to all bodies within the reach of our experiments, are to be esteemed the universal qualities of all bodies whatsoever.
4. In experimental philosophy we are to look upon propositions inferred by general induction from phenomena as accurately or very nearly true, notwithstanding any contrary hypotheses that may be imagined, till such time as other phenomena occur, by which they may either be made more accurate, or liable to exception.

Phenomena.

Six in number. 1. The satellites of Jupiter revolve round the planet in accordance with Kepler's Laws. 2. As do the satellites of Saturn. 3. Mercury, Venus, Mars, Jupiter and Saturn revolve round the Sun. 4. The periodic times of these five planets, and of the Earth about the

Sun, or of the Sun about the Earth, the fixed stars being at rest, are in accordance with Kepler's third Law. 5. The radii vectores of the primary planets do not describe equal areas in equal times about the Earth; but they do about the Sun. 6. As does the Moon about the Earth.

Newton summarizes the astronomical evidence for these statements.

The 42 propositions which make up the third book contains such propositions as that the satellites of Jupiter, the primary planets, and the moon are held in their orbits by an attraction varying according to the inverse square law; the attraction which holds the moon in her orbit is gravity; attraction is a property of all bodies; in moving towards the interior of a planet, the attraction is as the distance from the centre; the sun is perpetually in motion, but is never far from the C.G. of the whole system; the nodes and aphelia of the planets are fixed (corollaries and scholium dealing with perturbations); to find the axes, eccentricities, and aphelias; to find the proportions of the axis of a planet; on the variation of weight with latitude; lunar motion and inequalities; the tides; lunar theory; cometary theory. The two hypotheses which appear in Book III. are (1) the centre of the solar system is at rest, and (2) the effect of an equatorial ring of gravitating matter in causing precession is the same whether the ring be rigid or fluid (first demonstrated by Laplace).

The Scholium Generale which concludes the 'Principia' appeared first in the second edition. It is a statement of major importance for all who are interested in Newton's outlook. It is given here in full as translated by De Morgan.

Scholium Generale.

"The hypothesis of vortices is encumbered with many difficulties. Since a planet's radius describes areas proportional to the times, the periods of the parts of the vortex should be in the duplicate ratio of their distances from the sun. And since the planets' periods are in the sesquiplicate ratio of their distances, the periods of the parts of the vortex should also be in that ratio. Since the minor vortices about Jupiter, Saturn, etc., preserve their rotation, and swim quietly in the vortex of the sun, the periodic times of the parts of the solar vortex should be equal. The revolutions of the sun and planets about their axes, which should agree with the motions of the vortices, differ from all these proportions. The motion of comets is highly regular, follows the same laws as that of the planets, and cannot be explained by vortices. They are carried most excentrically to all parts of the heavens, which could not be unless the vortices were removed.

"Projectiles, in our atmosphere, feel only the resistance of the air; take that away, as in Boyle's vacuum, and the resistance ceases, since the light feather and the solid gold fall with the same velocity in this vacuum. And such is the case in the celestial spaces which are above the atmosphere of the earth. All bodies in those spaces must move perfectly freely; and hence the planets move perpetually, according to the laws explained, in orbits of given form and position. They will persevere in their orbits by the laws of gravity; but they could by no means originally have taken that regular orbital path by the same laws.

"The six principal planets revolve about the sun in circles concentric with the sun, in the same direction, and very nearly in the same plane. The ten moons revolve about the earth, Jupiter, and Saturn, in concentric circles, in the same direction, and very nearly in the plane of their planets' orbits. And all these regular motions have no origin in mechanical causes; for the comets are freely carried in very excentric orbits, and through all parts of the heavens. By such motion comets pass freely and easily through the planetary orbits, and in their aphelia, where they move more slowly and remain longer, they are very distant from each other, and attract each other very little. This most elegant group of sun, planets and comets can only arise from the design and government of a powerful and intelligent being. And if the fixed stars be centres of like systems, all these constructions of a similar design must be under the dominion of one being; the more since the light of the fixed stars is of the same nature as that of the sun, and all the systems give light to each other. And that the systems of the fixed stars may not fall into each other, he has placed them at immense distances from each other.

"He rules all things, not as the soul of the world, but as the lord of all things. And from his dominion, God the Ruler is often called *Παντοκράτωρ*. For God is a relative term, which refers to his servants; and deity is the rule of God, not over his own body, as those think who call him the soul of the world, but over his servants. The most high God is a being eternal, infinite, absolutely perfect; but a being however perfect, without dominion, is not God the Ruler. For we say, my God, your God, the God of Israel, the God of gods, the Ruler of rulers; but we do not say, my eternal one, your eternal one, the eternal one of Israel, the eternal one of gods; we do not say, my infinite one, or my perfect one. These appellations have no relation to servants. The word God always signifies ruler; but every ruler is not God. The rule of a spiritual being makes him God, true rule a true god, the highest rule the highest god, a feigned rule a feigned god. And from the reality of his dominion it follows that the true God has life, intellect, and power; and from his other perfections, that he is the highest or most perfect being. The Eternal is infinite, omnipotent, and omniscient; that is, endures from eternity to eternity, and is present from infinity to infinity; he rules all and knows all that is or can be. He is not eternity nor infinity, but eternal and infinite; he is not time nor space, but endures and is present. He endures for ever, and is everywhere present; and by his universal existence, both in time and space, he is the creator of time and space. Since any particle of space always is, and also some one indivisible moment of duration, certainly the framer and ruler of all things can never be non-existent at any time or in any place. Every thinking soul is, at different times, and in its different organs of sense or motion, the same indivisible person. There are parts of time successive, and parts of space co-existing; but neither in the person of man, that is, in his thinking principle, and much less in the thinking substance of God. Every man, considered as a sentient existence, is one and the same man throughout his whole life, and in every one of his organs of perception. God is one and the same God always and everywhere. He is omnipresent, not only virtually but substantially; for the first without the second cannot be. In him all things are contained and moved, but without mutual effect from each other. God no way feels the motion of bodies; they feel no resistance from the omnipresence of God. All men confess that the most high God exists necessarily, and by the same necessity always and everywhere. Whence also he is throughout alike, all eye, all ear, all brain, all arm, all thought, understanding and action, not in the human or corporeal manner, but in one altogether unknown. As a blind man has no idea of colour, so we have no idea of the modes in which the all-wise God perceives and understands all things. He is altogether without body or figure, and thence can neither be seen, heard, touched, nor worshipped under the likeness of any corporeal thing. We have ideas of his attributes, but we know nothing of the substance of any thing. We see only the figures and colours of bodies, we hear only sounds, we touch only external surface, we smell only odours, and taste tastes; we can perceive the inmost substance by no sense, by no reflected action; much less can we have an idea of the substance of God. Him we know only by his properties and attributes, and by the most wise and excellent structure and final causes of things; and we wonder at him for his perfections, but we venerate and worship him for his dominion. We worship him as servants, and a God without dominion, providence and final causes is nothing else but fate and nature. From a blind metaphysical necessity, which is certainly the same always and everywhere, no variation of things can arise. Every diversity of created things in space and in time can only arise from the ideas and the will of a necessarily existing being. But God is said allegorically to see, hear, speak, laugh, love, hate, desire, give, receive, rejoice, be made angry, fight, work, build, and construct. For every expression concerning God is drawn from human affairs by a certain figure of similitude, not perfect indeed, but something like. And this much of God, of whom it certainly belongs to natural philosophy to treat from phenomena.

" So far I have explained the phenomena of our heavens and the sea by the force of gravity; but I have not yet assigned the cause of gravity. Certainly this force arises from some cause, which penetrates even to the centre of the sun and planets, without diminution of its virtue; and which acts not according to the superficial magnitude of particles (as do many mechanical causes), but according to the quantity of solid matter; and of which the action is extended in every direction to immense distances, diminishing always in the duplicate ratio of the distances. Gravity towards the sun is compounded of the gravity towards all the several particles of the sun, and in receding from the sun diminishes accurately in the duplicate ratio of the distances, as far as the orbit of Saturn, as is made manifest by the quiescence of the aphelia of the planets, and even as far as the aphelia of the comets, if they also do not move. But I have not yet been able to deduce the reason of these properties of gravity from phenomena, and I do not frame hypotheses. For whatever is not deduced from phenomena, is to be called *hypothesis* : and hypotheses, whether metaphysical or physical, or of occult qualities, or mechanical, have no place in *experimental philosophy*. In this philosophy propositions are deduced from phenomena, and are rendered general by induction. So impenetrability, mobility, the impetus of bodies, and the laws of motion and gravity have become known. And it is enough that gravity really exists and acts according to laws explained by me, and suffices for all the motions of the heavenly bodies and of our sea.

" I might add something on that most subtile spirit which prevades solid bodies and lies hidden in them, by the force and action of which the particles of bodies attract each other at the smallest distances, and adhere when contiguous ; and by which electric bodies act at greater distances, both by attracting and repelling neighbouring particles ; and by which light is emitted, reflected, refracted, inflected, and gives heat to bodies ; and by which all sensation is excited, and the limbs of animals are moved at pleasure, namely, by its vibrations through the solid capillaments of the nerves, propagated from the external organs of sense to the brain, and from the brain to the muscles. But this could not be explained in few words ; nor is there a sufficient quantity of experiments to show and determine the laws according to which this *spiritus* acts."

The famous phrase " et hypotheses non fingo " (" and I do not frame hypotheses ") occurs in this scholium and has given rise to much discussion. One implication of the phrase may be noted here, as has been pointed out by Prof. E. T. Whittaker in his recent presidential address to the Royal Society of Edinburgh, that the physicist, in Newton's view, " is not under any obligation to concern himself with metaphysics ; he can give his undivided attention to investigating the laws which will enable him to predict phenomena, and can leave the deeper problems entirely out of account ; he can make it his purpose to describe rather than to explain." And indeed it is the root of that particular matter, as far as Newtonian discoveries are concerned. Ptolemy described the motions of the components of the solar system by means of a very complicated scheme ; Copernicus substituted for it a much simpler description, though he did not shake his description free from the concept of epicycles ; Newton subsumed the complicated motions of the solar system under a single formula of surpassing generality and simplicity but the *why* of the matter is as far away as ever.

The main purpose of this short essay is concerned with the ' Principia,' and the remaining years of Newton's life can receive but scant notice. He represented the University in Parliament from January 1689 to February 1690; he was seriously ill in 1692 and 1693; he was appointed Master of the Mint in 1697; in 1701 he resigned the Lucasian Chair; he again sat in Parliament over the period 1701–1705; he was elected President of the Royal Society in 1703, and held that office for the rest of his life; the first edition of his *Opticks* appeared in 1704; he was knighted by Queen Anne at a court held at Trinity Lodge in April 1705; the second edition of the 'Principia,' edited by R. Cotes, was published in 1713; the third edition, edited by H. Pemberton, was published in 1726. On March 20th, 1727, Newton died. His body lay in state on March 28th in the Jerusalem chamber, and was thereafter buried in Westminster Abbey.

SIBI GRATULENTUR MORTALES, TALE TANTUMQUE EXTITISSE
HUMANI GENERIS DECUS.

References.

Ball (R.), 'Great Astronomers,' 1895.
Ball (W. W. R.), 'Cambridge Papers,' 1918.
Ball (W. W. R.), Essay on Newton's 'Principia,' 1893.
Berry (A.), 'Short History of Astronomy,' 1898.
Bonnycastle (J.), Introduction to 'Astronomy,' 1811.
Brewster (D.), 'Memoirs of Sir Isaac Newton,' 1855.
Brodetsky (S.), 'Sir Isaac Newton,' 1927.
Buckley (H.), 'Short History of Physics,' 1927.
Cox (J.), 'Mechanics,' 1904.
Dampier-Whetham, 'History of Science,' 1930.
De Morgan, 'Budget of Paradoxes,' 2 vols., 1915. (Cites Greene and D'Israeli, I., p. 135.)
De Morgan, 'Penny Cyclopædia,' 1841. (Art. 'Principia.')
Encyc. Brit. 11th edition 1910–11. (Arts. "Figure of the Earth," "Navigation," "Newton.")
Hues (R.), 'Treatise on the Globes' (1594). (Hakluyt Soc. Edn., ed. C. R. Markham (1889).)
Libby (W.), 'Introduction to History of Science,' 1918.
Macpherson (H.), 'Makers of Astronomy,' 1933.
Newton (I.), 'Principia,' 2nd edition, 1713.
Newton (I.), 'Principia,' Glasgow issue of edn. Le Seur and Jacquier, 4 vols., 1823.
Sir Isaac Newton, 1727–1927. A Bi-centenary Evaluation of his Work, 1928.
Pemberton (H.), 'View of Sir Isaac Newton's Philosophy,' 1728.
Pledge (H. T.), 'Science since 1500,' 1939.
Singer (C.), 'Short History of Science,' 1941.
Stukeley (W.), 'Memoirs of Sir Isaac Newton's Life,' 1936.
Villamil (R. de), 'Newton: the Man,' 1931.
Whittaker (E. T.), Proc. R.S.E., ser. A, vol. lxi. pt. 3 (no. 20), 1942.

XC. *The Force required to give a Small Acceleration to a Slowly-moving Sphere carrying a Surface Charge of Electricity.*

By G. F. C. Searle, Sc.D., F.R.S.,
Cavendish Laboratory, Cambridge [*].

[Received August 28, 1942.]

§ 1. *Introduction.*

On the electromagnetic theory of the dynamics of electrified bodies, the rate of increase of the momentum of the ordinary matter in a non-magnetic and non-conducting body is the resultant of (1) the force applied to the body by strings or such-like means or by gravitation, and (2) the force exerted on the electrification by the electromagnetic field. The absence of conductivity ensures that no electric currents exist in the body and that the distribution of electricity on the body does not change. The force exerted by the field on a point-charge q, which moves with velocity $\mathbf{u}$, is made up of $q\mathbf{E}$ in the direction of $\mathbf{E}$, and of the vector product $q\mathrm{V}\mathbf{uB}$ perpendicular both to $\mathbf{u}$ and to $\mathbf{B}$, the magnetic induction. (Heavy type indicates vectors.) When a charged body moves, without rotation, parallel to a fixed straight line, the force typified by $q\mathrm{V}\mathbf{uB}$ is perpendicular to that line, and may be omitted in our reckoning of the force required to accelerate the body in the direction of the line.

The force required to accelerate in a fixed direction a uniformly charged sphere has been investigated in various ways, all of them subject to some restrictions. There is, of course, no unique solution, because the reaction of the radiation at any time t_0 depends not only upon the velocity, the acceleration and the successive time-derivatives of the acceleration of the sphere at t_0, but also upon the manner in which the velocity has changed in an interval previous to and ending at t_0. The problem can be brought to a manageable form by treating the motion as "quasi-stationary" and restricting the acceleration to be so small that the change of velocity which occurs in the time l/v, where l is the greatest distance between two points on the body and v is the velocity of light, is very small compared with v. At the other end of the scale, quite simple cases arise when the action is impulsive and the velocity of the body is suddenly changed from u_1 to u_2. In these latter cases, the force is finite and lasts for a finite time, provided u_1 and u_2 be less than v.

If a sphere with a surface charge q be suddenly started from rest with light speed, which is forcibly maintained, the sphere is always enclosed in its own wave system. The reaction of the radiation is a retarding force. The earliest statement, so far as I know, as to the reaction of the radiation on any moving system was made in 'Physical Review,' xi. p. 63 (July, 1900), in a review of Vol. II. of Oliver Heaviside's "Electromagnetic Theory," and gave the result for a sphere started at speed of light. The formula for the driving force F was printed as $q^2/2avt$ instead of $q^2/\mathrm{K}avt$; I did not see the printer's proof. This is the value of F when the sphere has travelled from rest for time t through a distance vt very great compared with its diameter. Details are given in "The Impulsive Motion of an Electrified Sphere," Phil. Mag. vol. xvii. p. 62 (Jan., 1909). Heaviside discusses the matter in 'Electromagnetic Theory,' vol. iii. pp. 122–125 *. Oliver and I had many friendly combats in those days of long ago.

---

[*] Communicated by the Author.

The exact results, obtained by Paul Hertz †, are as follows :—

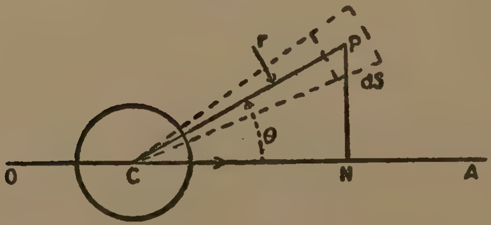
From $t=0$ to $t=a/v$,

$$\mathrm{F}=\frac{q^2}{2\mathrm{K}a^2}=\frac{\mu q^2 v^2}{2a^2}.$$

For $t=a/v$, and for greater values of t,

$$\mathrm{F}=\frac{q^2}{2\mathrm{K}a^2}\left\{\frac{2a}{vt}-\frac{a^2}{v^2t^2}\right\}.$$

Fig. 1.



The second result tends to $q^2/\mathrm{K}avt$ when vt/a becomes large.

Naturally, the theory of the quasi-stationary case is simplified when the *velocity* is restricted to be very small compared with that of light. When both the velocity and the acceleration approach the limit zero, the problem is easily solved, as appears below.

§ 2. *First Method.*

Let the centre of a sphere of radius a, which carries a uniformly distributed surface charge q, move along the straight line OA (fig. 1) with uniform acceleration. Let C be the position of the centre at time t, and let the velocity and the acceleration of C in the direction OA at time t be u and f. Since u is supposed very small compared with v, the velocity of light, and since $2af$, the change in the square of the velocity which occurs while C moves through a, is supposed very small compared with

---

* 'The Electrician,' Nov. 29, 1901.

† See G. F. C. Searle, Proc. Roy. Soc. A, lxxix. p. 552 (1907).

v^2, the electric field at time t at a fixed point near the sphere is, for a given position of C at that time, very little affected by the fact that C is not at rest but in motion. In other words, we may, to our approximation, consider that the sphere carries with it its statical field unchanged by the motion. Outside the sphere, this statical field is identical with that due to a point-charge q at C; inside the sphere, the statical field vanishes.

The moving electric field gives rise to a magnetic field. Let P be a fixed point, and PN the perpendicular from P on OA; let $\mathrm{CP}=r$, and $\mathrm{PCA}=\theta$. Then $\mathrm{PN}=r\sin\theta$. The solid angle subtended at C by the fixed circle of centre N, axis OA and radius NP, is $2\pi(1-\cos\theta)$. The flux of electric displacement from the charged sphere is distributed uniformly throughout the solid angle 4π, and hence the flux through the circle of radius NP is $(q/4\pi)2\pi(1-\cos\theta)$ or $\frac{1}{2}q(1-\cos\theta)$. The rate of increase of the flux through the circle, as C moves towards A, is $\frac{1}{2}q\sin\theta\,.\,d\theta/dt$. The velocity u of C has the component $u\cos\theta$ along CP towards P, and thus $dr/dt=-u\cos\theta$, a result needed later. The velocity u also has the component $u\sin\theta$ perpendicular to CP in the plane PCA, and hence, for $d\theta/dt$, the rate at which CP turns about P in this plane, we have $d\theta/dt=u\sin\theta/r$. The rate of increase of the flux now becomes $\frac{1}{2}qu\sin^2\theta/r$, and this is the electric current through the aperture of the circle. The line integral of the magnetic force H round the circle, of circumference $2\pi r\sin\theta$, equals 4π times this current. If B be the magnetic induction, so that $\mathrm{H}=\mathrm{B}/\mu$, we have

$$2\pi r\sin\theta\,.\,\mathrm{B}/\mu=4\pi\,.\,\tfrac{1}{2}qu\sin^2\theta/r,$$

or

$$\mathrm{B}=\mu qu\sin\theta/r^2, \quad \ldots\ldots\ldots \quad (1)$$

a well known result. In fig. 1, the direction of B at P is towards the reader.

As the sphere moves in the direction OA, the magnetic induction at P varies (1) because both r and θ change, and (2) because u is increasing at the rate $du/dt=f$. Hence

$$\frac{d\mathrm{B}}{dt}=\mu q\left\{\frac{\sin\theta}{r^2}\cdot\frac{du}{dt}+\frac{u\cos\theta}{r^2}\cdot\frac{d\theta}{dt}-\frac{2u\sin\theta}{r^3}\cdot\frac{dr}{dt}\right\}.$$

Since $d\theta/dt=u\sin\theta/r$, and $dr/dt=-u\cos\theta$, we have

$$\frac{d\mathrm{B}}{dt}=\mu q\sin\theta\left\{\frac{f}{r^2}+\frac{3u^2\cos\theta}{r^3}\right\}. \quad \ldots\ldots \quad (2)$$

A filament of changing magnetic induction, *i.e.*, a filament of magnetic current, in the fixed circle of radius NP, gives rise to electric force in planes containing OA. As u and f tend to zero, the electric force at C due to the filament of magnetic current tends to the value which it would have if $d\mathrm{B}/dt$ had a steady value. If there were a steady *electric* current i in the circle through P, the resultant *magnetic* force at C would be $2\pi i\mathrm{PN}^2/r^3$, or $2\pi i\sin^2\theta/r$, along CA. When there is a magnetic current in the circle, there is, relative to the case of the electric current, a change of sign and a factor $1/4\pi$. The magnetic current through the area $d\mathrm{S}$ (fig. 1) at P is

$dS \,.\, dB/dt$, and, as the figure is drawn, B at dS is towards the reader. Hence, if dE_1 be the electric force at C, in the direction of f, due to $dS \,.\, dB/dt$, we have

$$dE_1 = \left(\frac{-1}{4\pi}\right)\frac{2\pi\sin^2\theta}{r}\cdot\frac{dB}{dt}\,dS.$$

Hence, by (2), we have, with $dS = r\,dr\,d\theta$,

$$dE_1 = -\mu q\left\{\frac{f\sin^3\theta}{2r^2}+\frac{3u^2\sin^3\theta\cos\theta}{2r^3}\right\}dr\,d\theta.$$

For the resultant value of E_1 at C we have

$$E_1 = -\int_a^R\int_0^\pi \mu q\left\{\frac{f\sin^3\theta}{2r^2}+\frac{3u^2\sin^3\theta\cos\theta}{2r^3}\right\}dr\,d\theta, \quad . \quad . \quad . \quad (3)$$

where R is very large compared with a. The term involving u^2 vanishes on integration with respect to θ, and we find

$$E_1 = -\frac{2\mu qf}{3}\left(\frac{1}{a}-\frac{1}{R}\right).$$

The vast regions of "infinity" are the playground of waves started by the sphere in the dim and distant past, and thus integration "to infinity" is meaningless, unless conditions are strictly specified. But the effect of a spherical shell of thickness dr and of radius r is $-2\mu qf\,dr/3r^2$, and hence, when a is small enough, a moderate value of R may be treated as infinite for our purpose. We therefore put $R^{-1}=0$, and then obtain

$$E_1 = -2\mu qf/3a. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Under the restriction that u and f are very small, the electric force at points on the surface of the sphere due to the currents of magnetic induction may be treated as if those currents were permanent. Now, just as the magnetic force due to a filament of a steady electric current is identical with that due to the appropriate magnetic shell, so the electric force due to a filament of steady magnetic current is identical with that due to the appropriate electric shell. Such a shell is made up of elements of positive and of negative electricity. The force exerted on any element of charge dq' at any point outside the sphere by a charge q, evenly distributed over the surface of the sphere, is the same in direction and magnitude as that exerted on dq' by q at the centre of the sphere. Hence, the force exerted by dq' upon the sphere, when uniformly charged on its surface with quantity q, is the same as that exerted by dq' upon a charge q at the centre of the sphere. This result may be extended to the whole system of electric shells by which the whole system of magnetic currents can be replaced. Since all the magnetic currents lie outside the sphere, all the electric shells may be treated as lying outside the sphere.

We have found that the electric force at the centre C in the direction of f is given by $E_1 = -2\mu qf/3a$. The force exerted on a charge q at C would be $-2\mu q^2 f/3a$, and hence, if the reaction of the electromagnetic field on the charged sphere, in the direction of f, be X, we have

$$X = -\frac{2\mu q^2 f}{3a}. \quad \ldots \ldots \ldots \ldots \quad (5)$$

The mechanical force, F, which must be applied to the sphere in the direction of f to give it the acceleration f, is equal and opposite to X, and thus

$$F = 2\mu q^2 f/3a. \quad \ldots \ldots \ldots \ldots \quad (6)$$

The result is better expressed by saying that F/f tends to the limit $2\mu q^2/3a$, as both u and f tend to zero.

We can find E_1 at a point J on OA within the sphere at distance h from C towards A. Let $PJ = \rho$, and $PJA = \phi$. Then $\sin\phi = r\sin\theta/\rho$. At J

$$dE_1 = -\frac{1}{2}\frac{\sin^2\phi}{\rho}\cdot\frac{dB}{dt}\cdot dS = -\frac{1}{2}\frac{r^2\sin^2\theta}{\rho^3}\cdot\frac{dB}{dt}\cdot dS,$$

and hence, at J,

$$E_1 = -\frac{\mu q}{2}\int_a^R\int_0^\pi \frac{\sin^3\theta}{\rho^3}(fr + 3u^2\cos\theta)\,dr\,d\theta.$$

In the integral (3), the factor $\sin^3\theta\cos\theta$ ensures the vanishing of the part depending on u^2, and there was no need to take account of the magnitude of u. Actually, the u^2-term might have been neglected in (3) because it is of the second order in u, which tends to zero. For the latter reason it suffices to keep only the f-term, and then we have

$$E_1 = -\tfrac{1}{2}\mu q f\int_a^R\int_0^\pi \frac{\sin^3\theta\,d\theta}{\rho^3}\,r\,dr,$$

with $\rho^2 = r^2 + h^2 - 2rh\cos\theta$. Since J is inside and P is outside the sphere, both $r+h$ and $r-h$ are positive. With this in mind, we find

$$E_1 = -\frac{2\mu q f}{3}\int_a^R \frac{dr}{r^2} = -\frac{2\mu q f}{3}\left(\frac{1}{a} - \frac{1}{R}\right),$$

where we may put $R^{-1} = 0$. Hence E_1 has the same value at all points on CA inside the sphere.

All the electric shells are outside the sphere, and hence the electric force due to them at internal points is derivable from a potential. Since the electric field is symmetrical about OA and at points on the diameter, $2a$, lying along OA is directed along OA and, when f, u tend to zero, tends to a constant value, it follows that, throughout the sphere, the electric force is parallel to OA and E_1/f tends to the value $-2\mu q/3a$.

§ 3. *Second Method.*

The reaction of the radiation on a sphere with a surface charge can be found by direct integration, when the velocity and the acceleration of the sphere are both very small. For a volume charge, the integration might be troublesome.

If a point charge q pass at time t_0 through a fixed point O with velocity $\mathbf{u}$ and acceleration $\mathbf{f}$, the part depending upon $\mathbf{f}$ of the electric force at a fixed point P at distance r from O is, at time t_0+r/v, given by the vector equation *

$$\mathbf{E}=\frac{q}{\mathrm{K}r}\left[-\frac{\mathbf{f}}{(v-u\cos\theta)^2}+\frac{(v\mathbf{r}_1-\mathbf{u})f\cos\gamma}{(v-u\cos\theta)^3}\right]. \quad . \quad . \quad . \quad . \qquad (7)$$

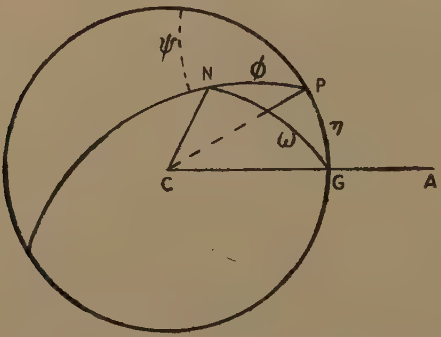
Here $\mathbf{r}_1$ is a unit vector in the direction from O towards P, and θ and γ are the angles made by $\mathbf{r}$ with $\mathbf{u}$ and $\mathbf{f}$.

In our case, u is negligible compared with v, and then

$$\mathbf{E}=\frac{q}{\mathrm{K}v^2r}\left[-\mathbf{f}+\mathbf{r}_1f\cos\gamma\right]=\frac{\mu q}{r}\left[-\mathbf{f}+\mathbf{r}_1f\cos\gamma\right], \quad . \quad . \quad . \quad . \qquad (8)$$

since $\mathrm{K}\mu v^2=1$. Hence, when u/v is very small, the electric force depending on $\mathbf{f}$ is independent of the direction and magnitude of u. By (8), $\mathbf{E}$ is the resultant of $(\mu q/r)f$ parallel to f, but in the opposite direction, and of $(\mu q/r)f\cos\gamma$ in the direction OP. If E_1 be the electric force at P in the direction of f, we find

Fig. 2.



$$\mathrm{E}_1=-\mu qf\sin^2\gamma/r. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (9)$$

In addition to the electric force, proportional to r^{-1} and to f, given by (7), there is, at t_0+r/v, an electric force radial from the point O′, which q would have reached at t_0+r/v, if it had continued to move with velocity u. This electric force depends upon u but not upon f; it is proportional to $\mathrm{O'P}^{-2}$, and so also to r^{-2}, since OO′ is proportional to r. A *system* of two point-charges moving along parallel lines with the same velocity may experience a *couple*, but it will not, as the result of this electric force, experience any *force* in any direction. For our present problem, we need not further consider this electric force.

Let C, the centre of the sphere, have acceleration f along CA, which meets the sphere in G (fig. 2). Let P be a point on the sphere, and let

---

* O. Heaviside, ‘Electromagnetic Theory,’ iii. p. 162; M. Abraham, *Theorie der Electrizität*, ii. p. 97 (1905); G. F. C. Searle, Phil. Mag., January, 1907, p. 148.

PCA=PCG=η. About P describe a sphere of radius r, where $r<2a$, let N be a point on the circle of intersection of the spheres, and let any radius of this circle subtend angle ϕ at C. Then NCP=ϕ, and r=NP=$2a \sin \frac{1}{2}\phi$. Let the plane NCP make angle ψ with the plane ACP. The projection of CP on CA or CG is $a \cos \eta$ and that of CN is $a \cos$ NCG or $a \cos \omega$, where $\cos \omega=\cos \phi \cos \eta-\sin \phi \sin \eta \cos \psi$. The projection of NP (*not* PN) on CA is $r \cos \gamma$, where γ is the angle between the *forward* directions of r and of f. The projection of NP on CA also equals $a \cos \eta$, the projection of CP on CA, *less* the projection of CN (*not* NC) on CA, and thus $r \cos \gamma=a \cos \eta-a \cos \omega$. Hence

$$2a \sin \tfrac{1}{2}\phi \cos \gamma=r \cos \gamma=a(\cos \eta-\cos \phi \cos \eta+\sin \phi \sin \eta \cos \psi)$$

$$=a(2 \sin^2 \tfrac{1}{2} \phi \cos \eta+2 \sin \tfrac{1}{2}\phi \cos \tfrac{1}{2}\phi \sin \eta \cos \psi).$$

Hence $$\cos \gamma= \sin \tfrac{1}{2}\phi \cos \eta+\cos \tfrac{1}{2}\phi \sin \eta \cos \psi,$$
and

$$\sin^2 \gamma=1- \sin^2 \tfrac{1}{2}\phi \cos^2 \eta- \cos^2 \tfrac{1}{2}\phi \sin^2 \eta \cos^2 \psi-\tfrac{1}{2} \sin \phi \sin 2\eta \cos \psi. \qquad (10)$$

Let N_0, P_0 be the positions of N, P at t_0, and let $N_0P_0=r_0$. A spherical wave due to the acceleration of an element of charge starts from N_0 at t_0 and reaches P later. The distance between N_0 and P, when the wave starting from N_0 reaches the moving point P, lies between $r_0/(1+u'/v)$ and $r_0/(1-u'/v)$, where u' is the greatest velocity of the sphere during the passage of the wave from N_0 to P. These two distances tend to equality with r_0 as u' tends to zero; ultimately we may neglect the displacement of the sphere during the passage of the wave from N_0 to P, and may disregard the distinction between N_0 and N. An element of area at N is $a^2 \sin \phi \, d\phi \, d\psi$, and the charge on it is $(q/4\pi) \sin \phi \, d\phi \, d\psi$. Hence, if E_1 be the component in the direction of f of the electric force at P due to the charge on the whole sphere, we have, by (9), since $r=2a \sin \frac{1}{2}\phi$,

$$E_1=-\frac{\mu q f}{4\pi}\int_0^\pi \frac{\sin \phi \, d\phi}{r} \int_0^{2\pi} \sin^2 \gamma d\psi=-\frac{\mu f}{4\pi a}\int_0^\pi \cos \tfrac{1}{2}\phi \, d\phi \int_0^{2\pi} \sin^2 \gamma d\psi.$$

By (10), $$\int_0^{2\pi} \sin^2 \gamma \, d\psi=2\pi(1- \sin^2 \tfrac{1}{2}\phi \cos^2 \eta-\tfrac{1}{2} \cos^2 \tfrac{1}{2}\phi \sin^2 \eta)$$

$$=\pi\{1+ \cos^2 \eta- \sin^2 \tfrac{1}{2}\phi(3 \cos^2 \eta-1)\}.$$

Hence

$$E_1=-\frac{\mu q f}{4a}\int_0^\pi \{1+ \cos^2 \eta-\sin^2 \tfrac{1}{2}\phi(3 \cos^2 \eta-1)\} \cos \tfrac{1}{2}\phi d\phi$$

$$=-\frac{\mu q f}{4a}\{2(1+\cos^2 \eta)-\tfrac{2}{3}(3 \cos^2 \eta-1)\}$$

$$=-2\mu q f/3a. \qquad (11)$$

Thus we have the notable result that E_1 is independent of η and is constant over the surface of the sphere.

If X be the force, in the direction of f, due to the action of the radiation on the charge, we have, as in (5),

$$X=-2\mu q^2 f/3a.$$

As in (6), $$F=2\mu q^2 f/3a.$$

The value of E_1 at C, the centre of the sphere, is easily found. If, as in fig. 2, N be any point on the sphere, the angle γ between the forward directions of NC and CA is $\pi-\omega$, and thus in (9) we replace $\sin^2\gamma$ by $\sin^2\omega$. The charge on the ring area defined by ω and $\omega+d\omega$ is $(q/4\pi a^2)2\pi a^2 \sin\omega\, d\omega$, or $\frac{1}{2}q\sin\omega\, d\omega$. Then, by (9), we have at C

$$E_1=-\mu f\int_0^\pi \frac{\sin^2\omega}{a}\cdot\tfrac{1}{2}q\sin\omega\, d\omega=-2\mu qf/3a, \quad . \quad . \quad . \quad (12)$$

as already found in § 2.

To examine the value of E_1 at points inside the sphere and on the axis CA, we take P on CA at distance h from C towards A. Then, if N be on the sphere and NCA$=\omega$, and if NP$=r$, the angle γ between r and f is $\pi-$NPA. Hence $\sin^2\gamma=a^2\sin^2\omega/r^2$, and thus at P

$$E_1=-\mu f\int_0^\pi \tfrac{1}{2}q\sin\omega\, d\omega\cdot\frac{a^2\sin^2\omega}{r^2}\cdot\frac{1}{r}$$

$$=-\tfrac{1}{2}\mu qfa^2\int_0^\pi \frac{\sin^3\omega\, d\omega}{(a^2+h^2-2ah\cos\omega)^{3/2}}.$$

In the integration we note that both $a+h$ and $a-h$ are positive. We find

$$E_1=-2\mu qf/3a.$$

Thus, the electric force E_1 has the same value at all points of the diameter which lies along CA.

When P is on the axis and outside the sphere, so that both $h+a$ and $h-a$ are positive, we find

$$E_1=-2\mu qfa^2/3h^3.$$

The approximations made in the electrical theory require than h/a is not large compared with unity.

§ 4. *Third Method.*

Let the centre C of the sphere be in steady motion along the axis Ox with velocity u_1 till the time $t=$, when the centre is at C_0, and afterwards let it move along Ox with varying velocity u. To produce this motion, a mechanical force F must be applied to the sphere. The work done by F from time $t=0$ to time $t=t$ is

$$\int_0^t Fu\, dt.$$

The work finds its equivalent in the gain of energy in the electromagnetic field.

At time t, when its centre is at C_t, let the sphere be suddenly arrested. The force required to hold it at rest does no work, since the sphere is at rest; this force is required only while the waves generated by the sudden stopping of the sphere or by its previous motion are passing over the sphere, and thus the force vanishes after the time $t+2a/v$. Since the motion from $t=0$ to $t=t$ was not uniform, electromagnetic disturbances travel outwards with velocity v, and the sudden stopping of the sphere at $t=t$ gives rise to a new disturbance superposed, to some extent, upon the disturbance which began before the time t.

Before $t=0$, the energy was U_1+T_1, where U_1 and T_1 are the electric and magnetic parts of the energy of the sphere when moving with uniform velocity u_1. When the compound wave has travelled out to infinity, the energy is U_0+W, where W is the energy in the wave system and U_0 is the electrostatic energy of the sphere at rest.

The gain in energy equals the work done by the force, and thus

$$\int_0^t Fu\,dt = U_0+W-U_1-T_1. \qquad (13)$$

Differentiating with regard to t, and noting that U_0, U_1, T_1 are independent of t, we obtain

$$Fu = dW/dt, \qquad (14)$$

and thus

$$F = u^{-1}dW/dt. \qquad (15)$$

By this method, due to Paul Hertz *, the problem of finding F is reduced to that of finding W as a function of t. I have extended Paul Hertz's method to other problems by considering the momentum carried off in the wave system.

In our case, the two waves, the first arising from the acceleration from $t=0$ to $t=t$, and the second from the sudden stopping at $t=t$, lie at $t=t'$ within the wave region which is bounded internally by a sphere of radius $v(t'-t)-a$, with C_t as centre, and externally by a sphere of radius $vt'+a$, with C_0 as centre. Ultimately, t' is supposed to become infinite. The wave due to the stopping lies between the spheres of radii $v(t'-t)-a$ and $v(t'-t)+a$ with C_t as centre. The disturbance due to the acceleration extends not only to the space between these latter spheres but also to the rest of the wave region.

The resultant electric and magnetic forces at any point of the wave region are $\mathbf{E}_f+\mathbf{E}_u$ and $\mathbf{H}_f+\mathbf{H}_u$, where $\mathbf{E}_f$, $\mathbf{E}_u$ are the (vector) electric forces in the waves due to the acceleration and to the stopping respectively, and similarly for $\mathbf{H}_f$, $\mathbf{H}_u$. Hence

$$W = \frac{1}{8\pi}\int\{K(\mathbf{E}_f+\mathbf{E}_u)^2+\mu(\mathbf{H}_f+\mathbf{H}_u)^2\}\,dv,$$

where the volume integral extends to the whole wave region. Of course, $\mathbf{E}_u=0$, $\mathbf{H}_u=0$, except in the volume between the spheres of radii $v(t'-t)-a$ and $v(t'-t)+a$ centred at C_t. Expanding the squares of the vectors, we have

$$W = W_u+W_f+W_{uf}, \qquad (16)$$

where, with $\mathbf{E}_u\mathbf{E}_f$, $\mathbf{H}_u\mathbf{H}_f$ as scalar products,

$$W_u = \frac{1}{8\pi}\int(KE_u^2+\mu H_u^2)\,dv, \qquad W_f = \frac{1}{8\pi}\int(KE_f^2+\mu H_f^2)\,dv,$$

$$W_{uf} = \frac{1}{4\pi}\int(K\mathbf{E}_u\mathbf{E}_f+\mu\mathbf{H}_u\mathbf{H}_f)\,dv.$$

---

* For references see G. F. C. Searle, Proc. Roy. Soc. A, lxxix. p. 550 (1907).

Here W_u is the energy in the wave given off when the sphere, previously moving with uniform velocity u, is suddenly stopped, W_f is the energy which would be found in the wave region if the sphere had moved with acceleration f till $t=t$ and had not been stopped at $t=t$. Further, W_{uf} is the energy due to the overlapping of the two disturbances. Then, since $du/dt=f$, we have

$$\frac{uF}{f}=\frac{dW_u}{du}+\frac{dW_f}{du}+\frac{dW_{uf}}{du}.$$

It is obvious that, as f tends to zero, the second and third terms tend to zero. Hence, for infinitesimal accelerations, we have

$$uF/f=dW_u/dt,$$

or

$$F=\frac{f}{u}\frac{dW_u}{dt}. \qquad \qquad (17)$$

When u/v is not very small, detailed calculation is required to find W_u; the work was first done by Oliver Heaviside. When, however, u/v is infinitesimal, simple methods suffice. When u/v is small, the magnetic energy of the sphere in steady motion is $\mu q^2u^2/3a$, and, as far as u^2, the electric energy is identical with the electrostatic energy, viz., $q^2/2Ka$ or $\mu q^2v^2/2a$. The total energy is $\mu q^2a^{-1}(\frac{1}{2}v^2+\frac{1}{3}u^2.)$ After the sudden stopping of the sphere, the total energy ultimately becomes $\frac{1}{2}\mu q^2a^{-1}v^2+W_u$. The total is unchanged, because the stopping force does no work. Hence

$$W_u=\mu q^2u^2/3a.$$

Hence, by (17),

$$F=\frac{f}{u}\frac{dW_u}{du}=\frac{2\mu q^2f}{3a},$$

as we have already found.

§ 5. *Finite Velocity and Finite Acceleration.*

When u is finite but less than v, and when f is constant and $2af$ is small compared with $v(v-u)$, we have, as far as f^2,

$$F=m_lf-\frac{2\mu q^2v^3uf^2.}{(v^2-u^2)^3}, \qquad \qquad (18)$$

where m_l, the "longitudinal electromagnetic mass" of the charged sphere, is given by

$$m_l=\frac{\mu q^2}{2a}\left\{\frac{2v^4}{u^2(v^2-u^2)}-\frac{v^3}{u^3}\log\frac{v+u}{v-u}\right\}.$$

In series form,

$$m_l=\frac{\mu q^2}{a}\left\{\frac{2}{3}+\frac{4u^2}{5v}+\frac{4u^6}{v7}+\ldots\right\}.$$

The term in (18) involving f^2 is independent of the radius of the sphere.

A. Sommerfeld * found an expression for the mechanical force F applied to the sphere at time t, when it moves with uniform acceleration f

---

* *Zur Eletronentheorie, II. Göttinger Nachrichten*, 1904, p. 410.

and at time t, has velocity u. The work is elaborate and his expression for F is complicated. When it is expanded in powers of f as far as f^2, the formula for F agrees with (18). Sommerfeld did not carry the expansion beyond the first power of f. In an essay (unpublished) submitted for the Adams prize (1908), I applied Paul Hertz's method and, going as far as f^2, obtained the result given in (18). I also found that, when f is variable, a term

$$-\frac{2\mu q^2 v^3}{3(v^2-u^2)^2}\cdot\frac{df}{dt}$$

must be added to the right side of (18).

In the essay I expanded Sommerfeld's expression as far as f^4 and found that

$$\frac{4\mu q^2 v^4 f^3 a(v^2+7u^2)}{3(v^2-u^2)^5}-\frac{2\mu q^2 v^3 u f^4 a^2(27v^4+86v^2u^2+7u^4)}{3(v^2-u^2)^7}$$

should be added to the right side of (18).

XCI. *Note on Einstein's Theory of Gravitation.*

By R. A. Houstoun, D.Sc., University of Glasgow *.

[Received July 16, 1942.]

As is well known, Einstein's theory of gravitation predicts three small effects not previously known. Two of these are the deflexion of light rays passing near the sun and the advance of Mercury's perihelion. I have found that these are given by Fermat's principle and the principle of least action respectively.

Einstein obtains the following differential equation for the path of Mercury :

$$\frac{d^2u}{d\theta^2}+u=\frac{m}{h^2}+3mu^2.$$

If the m/h^2 term is omitted, we get the path of the light ray. If the $3mu^2$ term is omitted, we obtain the motion in an ellipse given by Newton's theory of gravitation. The action takes place in a plane through the sun and we can therefore restrict ourselves to two-dimensional polar co-ordinates.

Applying Fermat's principle of least, or rather of stationary time, we have for a ray in an isotropic medium

$$\delta\int\sqrt{1+\left(r\frac{d\theta}{dr}\right)^2}\,dr=0. \qquad (1)$$

---

* Communicated by the Author.

This leads to
$$\frac{d^2u}{d\theta^2}+u=0$$
and the straight line.

Applying the principle of least action, we have for a planet in the field of the sun

$$\delta\int\sqrt{\frac{2km}{r}-\mathrm{C}}\sqrt{1+\left(r\frac{d\theta}{dr}\right)^2}\,dr=0, \quad . \quad . \quad . \quad . \quad . \quad . \tag{2}$$

where k is the gravitational constant, m is the mass of the sun and C is a constant. This leads to

$$\frac{d^2u}{d\theta^2}+u=\frac{km}{h^2},$$

the difference from the equation above being due to the fact that Einstein uses special units.

Let us return to Fermat's principle and assume that the medium is very slightly ælotropic, writing

$$\sqrt{n_r{}^2+\left(n_\theta r\frac{d\theta}{dr}\right)^2}\,dr$$

for the element of optical length, where

$$n_r=n_\theta{}^2=\left(1-\frac{2km}{rc^2}\right)^{-1}$$

and c is the velocity of light *in vacuo*. The application of Fermat's principle then becomes

$$\delta\int\sqrt{n_r{}^2+\left(n_\theta r\frac{d\theta}{dr}\right)^2}\,dr=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \tag{3}$$

and the application of least action

$$\delta\int\sqrt{\frac{2km}{r}-\mathrm{C}}\sqrt{n_r{}^2+\left(n_\theta r\frac{d\theta}{dr}\right)^2}=0. \quad . \quad . \quad . \quad . \quad . \quad . \tag{4}$$

These two equations give respectively *exactly* the same paths as Einstein obtains for the passage of the light ray and the planet in the sun's gravitational field.

The expression used above for optical length is the same as is used in dealing, for example, with the extraordinary ray in Iceland spar. The novel feature here is the numerical values of n_r and n_θ. n_r is only $1+8{\cdot}6\times10^{-6}$ at the surface of the sun, where it is greatest. It is this very slight ælotropy of the medium which constitutes the difference between Newton and Einstein.

I give the calculation for the case of equation (4): (1), (2) and (3) can be obtained from (4) by omitting the appropriate terms.

The formula is $d\mathrm{V}/d\dot{y}=$const., where V is the integrand and $\dot{y}=d\theta/dr$. Writing h for the constant, we obtain in this case

$$\sqrt{\frac{2km}{r}-\mathrm{C}}\,\frac{n_\theta{}^2r^2\dfrac{d\theta}{dr}}{\sqrt{n_r{}^2+\left(n_\theta r\dfrac{d\theta}{dr}\right)^2}}=h.$$

At an apse where $d\theta/dr$ is infinite, h becomes $n_\theta rv$, and thus, when $n_\theta=1$, has its usual meaning of twice the areal velocity. This equation becomes

$$\left(\frac{2km}{r}-\mathrm{C}\right)n_\theta{}^4r^4\left(\frac{d\theta}{dr}\right)^2=h^2\left(n_r{}^2+n_\theta{}^2r^2\left(\frac{d\theta}{dr}\right)^2\right),$$

and on changing r to u and substituting for n_r and n_θ,

$$\left(\frac{du}{d\theta}\right)^2+u^2=\frac{2kmu-\mathrm{C}}{h^2}+\frac{2kmu^3}{c^2}.$$

This becomes

$$\frac{d^2u}{d\theta^2}+u=\frac{km}{h^2}+\frac{3kmu^2}{c^2},$$

the difference from Einstein's equation being again due to his special units.

The question arises as to the meaning of these equations, whether they are merely a mathematical curiosity or afford a means of obtaining Einstein's results without using tensors. I hold this second view.

Einstein's third gravitational effect is the "red shift," the displacement towards the red of a spectral line emitted by an atom vibrating in a gravitational field. Owing to the time dilatation a spectral line emitted by an atom travelling with a high velocity is also displaced towards the red. This has been proved directly by the Ives–Stilwell experiment on canal rays. The formulæ for the two effects agree if the velocity in the second case is the velocity acquired by falling from infinity in the first case. This coincidence is undoubtedly significant; it has been noted by H. B. Phillips (1) and H. E. Ives (2).

Ives assumes that an interferometer stationary in a gravitational field assumes dimensions and clock rate as though in motion with the equivalent velocity. From this he deduces the correct expression for the velocity of light in a gravitational field. The red shift follows automatically from his assumption. He does not work out the expression for the star shift nor make any application to the motion of a planet.

Let $\kappa=(1-v^2/c^2)^{-\frac{1}{2}}$, where v is the velocity from infinity. $v^2=2km/r$. Let T′ be the time required by light to travel a distance L′ in the gravitational field. Then the velocity of light is given by $\mathrm{L}'=c\mathrm{T}'$. Let T and L be the same quantities measured by a metre-stick and clock unaffected by the field. In the radial direction $\mathrm{L}'=\kappa\mathrm{L}$, $\mathrm{T}'=\mathrm{T}/\kappa$, the velocity

$$c_r=\frac{\mathrm{L}}{\mathrm{T}}=\frac{\mathrm{L}'}{\mathrm{T}'\kappa^2}=\frac{c}{\kappa^2},$$

and the refractive index $n_r=\kappa^2$. In the transverse direction $\mathrm{L}'=\mathrm{L}$, the velocity

$$c_\theta=\frac{\mathrm{L}}{\mathrm{T}}=\frac{\mathrm{L}'}{\mathrm{T}'\kappa}=\frac{c}{\kappa},$$

and the refractive index $n_\theta=\kappa$. These are precisely the values which I have adopted earlier in this paper.

The equivalence between the gravitational field and system in motion does not extend to the Doppler effect. The field produces only the red shift and no Doppler effect in addition. This is to be expected, for there is a distinction in kind between the two effects ; one is due to the motion of the source, the other to changes in dimensions and time rate caused by this motion.

Ives does not state why the gravitational field should be equivalent to the one particular velocity. We must here fall back on Einstein's principle of equivalence, according to which an observer falling freely from infinity would think that dimensions and clock rate are constant. He is unaware that he is being accelerated, this being the nature of gravitation. Let us consider, therefore, two observers A and B situated at a great distance from the sun. Let A be at rest relatively to the sun and let B start to fall towards it. Dimensions and clock rate must be constant for B ; hence they appear to A, as if he were falling *away* from the sun with a velocity equal and opposite to that of B. This is all there is to it.

Thus I assume Einstein's principle of equivalence and also the fact that space and time are distorted in a gravitational field. But I do not assume that $\delta\int ds=0$ for the path of a particle or that $ds=0$ for the path of a ray ; these principles are replaced by least action and least time. Einstein's stationary principles are, of course, sufficient, but I do not regard them as necessary, and in any case they add greatly to the mathematical difficulty of the subject. They have also the disadvantage of abolishing the concept of force.

H. B. Phillips uses the principle of equivalence in what is practically the same way as I do, though he gives a mathematical explanation. He rejects the general principle of relativity, namely that the laws of physics can be stated in a form that is the same for all reference systems, and he derives the correct expressions for the star shift and the advance of Mercury's perihelion by an application of the restricted principle of relativity. Consequently he does not use the concept of force. I did not become aware of his paper until my own was written.

Surprisingly little has been written on Einstein's theory of gravitation from an independent standpoint. The only papers known to me in addition to the two already cited are those of R. A. M. Kearney (3) and A. Eagle (4). R. A. M. Kearney obtains Einstein's three effects by an original use of the curvature of space, and A. Eagle explains the deflexion of the ray by supposing that the index of refraction of the ether varies with the distance from the centre of the sun. It is possible to get the correct result for the deflexion in this way, but as he uses an isotropic medium, he does not obtain the same path as Einstein.

Note on Relativistic Mass.

It is also simpler, I think, to base the explanation of the relativistic change of mass on the FL contraction and time dilatation, which are physical concepts and easier to grasp than the Lorentz transformation. Let us suppose that the deflecting field is provided by a point charge at a great distance and that the electron is projected, so that it cuts the lines of force at right-angles with velocity v. As the laws of dynamics hold only for small relative motion of observer and electron, we must assume an observer near the electron travelling with almost the same velocity as the latter. We must also assume the existence of a stationary observer who thinks that the other's clock is going slow, and that the other's metre-stick contracts when it is pointed in the direction of the relative motion, and naturally tries to correct for the defects of his colleague's measuring instruments.

As before, let $\kappa=(1-v^2/c^2)^{-\frac{1}{2}}$. Let t, E be the time and field intensity for the stationary observer and t', E′ the same quantities for the moving observer. The transverse displacement y has the same value for both observers. The stationary observer writes

$$\frac{d^2y}{dt^2}=\frac{1}{\kappa^2}\frac{d^2y}{dt'^2}$$

to allow for the other's clock going slow.

In order to find E we draw a sphere through the electron with its centre at the point charge. The charge and consequently the induction through the sphere is, like the velocity of light, the same for both observers. This equivalence holds for every element of the surface. But to the stationary observer the area of the moving observer's element appears to contract, the lines of force come closer together and the field intensity appears to increase. Thus $\mathrm{E}'=\kappa\mathrm{E}$.

Now
$$m_0\frac{d^2y}{dt'^2}=e\mathrm{E}'.$$

On making the substitutions

$$m_0\kappa^2\frac{d^2y}{dt^2}=e\kappa\mathrm{E},\quad m_0\kappa\frac{d^2y}{dt^2}=e\mathrm{E},\quad \text{and}\quad m=m_0\kappa.$$

References.

(1) J. Math. Phys. i. pp. 177–190 (1922).
(2) J. Opt. Soc. of Amer. xxix. pp. 183–187 (1939).
(3) Phil. Mag. xxviii. pp. 694–701 (1939).
(4) Math. Gaz. xviii. pp. 145–164 (1934).

XCII. *The Determination of High-Frequency Transmission Line Constants by Measurements at Special Frequencies.*

By J. C. SIMMONDS *.

[Received August 24, 1942.]

IT is well-known that the attenuation constant and the characteristic impedance of a high-frequency transmission line, as well as the primary constants and certain other quantities, can, in general, be calculated from the open-circuit and closed-circuit input impedances of the line. At certain frequencies the input impedances of the line become pure resistances, and from these values it is not possible to calculate the primary constants. However, the attenuation constant and the characteristic impedance can be obtained from these resistance values, whilst from a knowledge of the frequency the phase constant can be found.

A great advantage of working with these special frequencies is that only simple calculations are involved.

The frequencies at which the input impedances of the line are purely resistive are usually known as " resonance frequencies " and are such that the electrical length of the line is a multiple of one-quarter of a wave-length. At any resonance frequency one of the input resistances is greater than and one is less than the characteristic impedance. If these values of resistance are denoted by R_{max} and R_{min}, respectively, then it can be shown that the attenuation of the line is given by

$$\alpha l = \tanh^{-1} \sqrt{\frac{R_{min}}{R_{max}}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where α is the attenuation constant and l is the length of the line, and that the characteristic impedance is given by

$$Z_0 = \sqrt{R_{min} \cdot R_{max}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Since the length of the line at the resonance frequences is a multiple of a quarter wave-length, $\beta l = \frac{n\pi}{2}$ at these frequencies, β being the phase constant and n being any positive integer. This information, together with a knowledge of the approximate velocity of phase propagation, is sufficient to enable the phase constant to be calculated, and from this the exact value of the velocity of phase propagation can be found.

In the derivation of equations (1) and (2) it is assumed that the characteristic impedance has no reactive component. Although this assumption is probably well justified for low-loss lines at high frequencies, it is not always justified, particularly at comparatively low frequencies. It is now proposed to investigate, in detail, the effect of the reactive component of the characteristic impedance.

---

* Communicated by the Author.

The input impedance of a transmission line of characteristic impedance Z_0, propagation constant P and length l, when terminated in an impedance Z_R is given by

$$Z=Z_0\left\{\frac{Z_R \cosh Pl+Z_0 \sinh Pl}{Z_0 \cosh Pl+Z_R \sinh Pl}\right\}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

When the far-end of the line is close-circuited, Z_R is zero and equation (3) reduces to

$$Z_c=Z_0 \tanh Pl, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

whilst an open-circuit at the far-end corresponds to an infinite Z_R and gives

$$Z_f=Z_0 \coth Pl. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Now let

$$P=\alpha+j\beta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

then, if equation (6) is substituted in equations (4) and (5) and the quotient and the product of these quantities taken, equations (1) and (2) result, providing Z_0 is purely resistive and the electrical length of the line is such that

$$\beta l=\frac{n\pi}{2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

By making use of equation (6), equations (4) and (5) can be written in the form

$$Z_c=Z_0\left\{\frac{\sinh 2\alpha l+j \sin 2\beta l}{\cosh 2\alpha l+\cos 2\beta l}\right\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

$$Z_f=Z_0\left\{\frac{\sinh 2\alpha l-j \sin 2\beta l}{\cosh 2\alpha l-\cos 2\beta l}\right\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

and the reactive component of the characteristic impedance can be taken into account by writing

$$Z_0=R_0+jX_0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

The condition that the imaginary component of equation (8) be zero is then seen to be

$$R_0 \sin 2\beta l+X_0 \sinh 2\alpha l=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and the similar condition for equation (9) is

$$R_0 \sin 2\beta l-X_0 \sinh 2\alpha l=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

It will be observed that these equations reduce to equation (7) if either X_0 or αl is zero. In general, however, the reactive component of the characteristic impedance causes the corresponding open-circuit and closed-circuit resonance frequencies to differ.

Effect of X_0 on the Resonance Frequencies.

An approximate estimate of the amount the resonance frequencies differ from the frequencies which make the electrical line length a multiple of a quarter wave-length may be made by assuming extreme values for the variables of equations (11) and (12). For this purpose let $2\alpha l$ at the first resonance frequency $\left(\beta l=\frac{\pi}{2}\right)$ be unity, a value which would rarely be exceeded in practice, and let $\frac{X_0}{R_0}=a=-\frac{1}{10}$. In these circumstances, it is seen that the first closed-circuit resonance frequency is decreased by some 4 per cent., whilst the first open-circuit resonance frequency is increased by a similar amount.

Equations (11) and (12) show that, providing it is permissible to neglect the variation of α and a with small changes of frequency, the open-circuit and closed-circuit resonance frequencies are symmetrically disposed, on a linear scale, about that frequency which makes the electrical line length a multiple of one-quarter wave-length. To determine the effect of variation of α and a with frequency, let the value of α at the frequency corresponding to $\beta l=\frac{n\pi}{2}$ be α_n and the value of a be a_n, and let the frequency be f_n. The attenuation of lines which have negligible dielectric loss varies as the square root of the frequency, whilst if it were possible to have a line with only dielectric loss the attenuation would vary directly as the frequency. Thus, at a frequency slightly different from f_n, say, $f_n(1+p)$, the attenuation will be $\alpha_n(1+\varDelta)$, where $\varDelta$ lies between p and $p/2$ depending upon the relative values of the conductor loss and the dielectric loss. Similarly, at the frequency $f_n(1+p)$, the value of a may be written $a_n(1-\delta)$, where δ is approximately equal to $p/2$ because, for small variations, a varies approximately as the inverse square root of the frequency. (The special case when $a=0$ will not be considered.) Now, let β_1, p_1, δ_1 and $\varDelta_1$ correspond to a resonance frequency slightly greater than f_n and β_2, p_2, δ_2 and $\varDelta_2$ correspond to a resonance frequency slightly less than f_n. Then, if $a \sinh 2\alpha l$ is small,

$$2\beta_1 l=v\pi+a_n(1-\delta_1)\sinh 2\alpha_n l(1+\varDelta_1), \quad . \quad . \quad . \quad . \quad (13)$$

and

$$2\beta_2 l=v\pi-a_n(1+\delta_2)\sinh 2\alpha_n l(1-\varDelta_2), \quad . \quad . \quad . \quad . \quad (14)$$

where v is a positive, odd integer.

The mean value of β is, therefore, given by

$$\beta_m l=\frac{v\pi}{2}+\frac{a_n}{4}\{(1-\delta_1)\sinh 2\alpha_n l(1+\varDelta_1)-(1+\delta_2)\sinh 2\alpha_n l(1-\varDelta_2)\}, \quad (15)$$

and $\beta_m l$ is seen to differ from a multiple of $\pi/2$ by an amount very closely equal to

$$+\frac{a_n}{4}\,(\varDelta_1+\varDelta_2)2\alpha_n l\cosh 2\alpha_n l-(\delta_1+\delta_2)\sinh 2\alpha_n l\}, \quad . \quad . \quad (16)$$

or by an amount between

$$\frac{a_n(p_1+p_2)}{8}\{2\alpha_n l\cosh 2\alpha_n l-\sinh 2\alpha_n l\} \quad . \quad . \quad . \quad . \quad (17)$$

and

$$\frac{a_n(p_1+p_2)}{4}\left\{2\alpha_n l\cosh 2\alpha_n l-\frac{\sinh 2\alpha_n l}{2}\right\} \quad . \quad . \quad . \quad . \quad (18)$$

Similar expressions can be found when v is an even, positive integer. Therefore, if $\alpha_n l$ is never allowed to exceed 1·0 neper this quantity is always less than

$$\frac{3a_n \partial f}{2}, \qquad (19)$$

where ∂f is the fractional difference between the two corresponding resonance frequencies. From equations (11) and (12), it can be shown that

$$a_n \approx \frac{\partial f n\pi}{2 \sinh 2\alpha_n l}. \qquad (20)$$

Therefore, the fractional amount by which $\beta_m l$ differs from $n\pi/2$ is less than

$$\frac{(\partial f)^2}{4}\{2\alpha_n l \coth 2\alpha_n l - \tfrac{1}{2}\}, \qquad (21)$$

and for the extreme values of $2\alpha l$ and a assumed above is less than 0·2 per cent. at the first resonance frequency, *i. e.*, $n=1$. Thus, in most practical cases, the frequency which makes βl a multiple of $\pi/2$ can be found by taking the mean value of the two resonance frequencies. For the purpose of computing the quantity given by equation (21) it is permissible, of course, to use the value of $\coth 2\alpha l$ calculated from equation (1) in place of $\coth 2\alpha_n l$.

Effect of X_0 *on the Expression for the Characteristic Impedance.*

By substituting the conditions for zero reactance given by equations (11) and (12) in equations (8) and (9) respectively, the closed-circuit and the open-circuit input resistances, R_c and R_f, at the resonance frequencies may be obtained. Taking the product of R_c and R_f and writing a and α in terms of a_n and α_n as in the previous section, gives, after some manipulation and approximation

$$R_c R_f = R_0^2\left\{1 + a_n^2 - \frac{2\alpha_n l}{\sinh 2\alpha_n l} \cdot \partial f\right\}. \qquad (22)$$

In the derivation of this expression $(\Delta_1 + \Delta_2)$ has been assumed equal to ∂f; actually it is less than ∂f, unless the loss of the feeder under consideration is directly proportional to the frequency, and has a minimum value of $\partial f/2$. Using equation (20), a_n can be eliminated from equation (22) to give

$$R_c R_f = R_0^2\left\{1 + \left(\frac{n\pi \partial f}{2 \sinh 2\alpha_n l}\right)^2 - \frac{2\alpha_n l \partial f}{\sinh 2\alpha_n l}\right\}. \qquad (23)$$

The error introduced by employing equation (2) to calculate the characteristic impedance is, therefore,

$$\frac{1}{2}\left\{\left(\frac{n\pi \partial f}{2 \sinh 2\alpha_n l}\right)^2 - \frac{2\alpha_n l \partial f}{\sinh 2\alpha_n l}\right\}. \qquad (24)$$

In a similar manner, when the quotient of R_c and R_f is taken it can be shown that

$$\frac{R_{\min}}{R_{\max}} = \tanh^2 \alpha_n l\left\{1 + \frac{2\alpha_n l(\Delta_1 - \Delta_2)}{\sinh 2\alpha_n l} + \left(\frac{n\pi \partial f}{2 \sinh 2\alpha_n l}\right)^2 \cosh 2\alpha_n l\right\}. \qquad (25)$$

The error introduced by employing equation (1) to calculate the attenuation is, therefore,

$$\frac{1}{2}\left\{(\Delta_1-\Delta_2)\cdot\frac{2\alpha_n l}{\sinh 2\alpha_n l}+\left(\frac{n\pi\,\partial f}{2\sinh 2\alpha_n l}\right)^2\cosh 2\alpha_n l\right\}. \quad . \quad . \quad (26)$$

The order of $(\Delta_1-\Delta_2)$ can be estimated from equation (21), but generally the term involving this factor will be quite negligible.

Both expression (24) and expression (26) are only approximations. They could be made more exact by considering the relative contributions of the dielectric and the conductor to the attenuation, but this would greatly reduce their practical value. However, these expressions can be used to obtain the order of the error (for this purpose the value of αl calculated from equation (1) may be taken) and, if there appears to be any danger of the errors exceeding permissible values, the measuring technique can be modified accordingly.

Errors Introduced by Measuring the Open- and Closed-Circuit Impedances at the same Frequency.

Assume that the closed-circuit input impedance at a closed-circuit resonance frequency has been determined. The reactive component is zero, of course, and from equations (8) and (11) the resistive component is seen to be

$$R_c=R_0\left\{\frac{(1+a^2)\sinh 2\alpha l}{\cosh 2\alpha l+\sqrt{1-a^2\sinh^2 2\alpha l}}\right\}. \quad . \quad . \quad . \quad . \quad (27)$$

The resistive component of the open-circuit input impedance at this same frequency is, from equations (9) and (11), given by

$$R_f=R_0\left\{\frac{(1-a^2)\sinh 2\alpha l}{\cosh 2\alpha l-\sqrt{1-a^2\sinh^2 2\alpha l}}\right\}, \quad . \quad . \quad . \quad . \quad (28)$$

whilst the reactive component is given by

$$X_f=\frac{2X_0\sinh 2\alpha l}{\cosh 2\alpha l-\sqrt{1-a^2\sinh^2 2\alpha l}}. \quad . \quad . \quad . \quad . \quad . \quad (29)$$

It follows from equations (27) and (28) that approximately

$$R_0=\sqrt{R_cR_f}\left(1+\frac{a^2}{2}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

and

$$\tanh\alpha l=\sqrt{\frac{R_c}{R_f}}\left\{1-\frac{a^2}{2}(2+\cosh 2\alpha l)\right\}. \quad . \quad . \quad . \quad . \quad (31)$$

The latter equation applies at any "even" resonance frequency, but a similar expression can be found for any "odd" resonance frequency.

Therefore, in general,

$$\tanh\alpha l=\sqrt{\frac{R_{\min}}{R_{\max}}}\left\{1-\frac{a^2}{2}(2+\cosh 2\alpha l)\right\}. \quad . \quad . \quad . \quad . \quad (32)$$

It is interesting to note that the attenuation does not appear in equation (30), and that the error introduced by using the simple formula for Z_0

is smaller than that introduced when the method of the previous section is employed. If it is possible to measure X_f, the open-circuit reactance, the value of a can be found, for

$$\frac{X_f}{R_f}=\frac{2a}{1-a^2}=r \text{ say}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

as may be seen from equations (28) and (29), and thence

$$a \approx r/2.$$

This result is also obtained for $\frac{X_c}{R_c}$, when measured at an open-circuit resonance frequency. Thus, it is possible to determine a and thence to estimate the errors introduced by employing equations (1) and (2) for the calculation of the attenuation and the characteristic impedance. This method appears to offer the practical advantage that only one resonance frequency need be found and, providing the reactive component can be measured, the true value of R_0 can be found and a very close approximation to α can be made.

Theoretically it is possible to calculate the other constants of the feeder from the data obtained. This may, however, prove impracticable due to the difficulty of determining the resonance frequency with sufficient accuracy and/or due to the difficulty of measuring the reactive component.

Measurements at Odd Multiples of One-eighth of a Wave-length.

For the purpose of this section Z_0 will be assumed to be entirely resistive and βl will be assumed equal to an odd multiple of $\pi/4$. In these circumstances, equations (8) and (9) reduce to

$$Z_c=\frac{R_0}{\cosh 2\alpha l}(\sinh 2\alpha l \pm j), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

$$Z_f=\frac{R_0}{\cosh 2\alpha l}(\sinh 2\alpha l \mp j), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

and show that, at those frequencies which correspond to a line length of an odd multiple of one-eighth of a wave-length, the magnitudes of the open- and closed-circuit impedances are both equal to the characteristic impedance of the line. Further,

$$R_c=R_f=R_0 \tanh 2\alpha l, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

and

$$\pm X_c=\mp X_f=R_0 \operatorname{sech} 2\alpha l, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

whence

$$\frac{R_c}{|X_c|}=\frac{R_f}{|X_f|}=\sinh 2\alpha l. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

An advantage of making measurements at these special frequencies is that magnitudes of the impedances are equal. Equation (38) may be of use at high frequencies, since at such frequencies it is often much easier to measure reactances than resistances. It will be observed that the calculations are quite as simple as those involved in resonance frequency measurements.

XCIII. *Supplementary Note on Newtonian Mass, its Definition and Determination.*

By GLEN H. DRAPER*.

[Received August 14, 1942.]

GIVEN $\ddot{\chi}_i = X(t)$, the accelerations of each of n bodies ($i=1, 2, \ldots n$) as measured from any origin, either on one of the bodies or from an abstract point in space.

The system of n bodies is " an isolated Newtonian system " if there exist n definitive constants m_i which satisfy identically the $3(s+1)$ equations

$$\frac{d^s}{d^s t}\left(\sum_{j=1}^{n} m_j \sum_{i=1}^{n} m_i(\ddot{\chi}_i - \ddot{\chi}_j)\right) = \sum_{j=1}^{n} m_j M \frac{d^s}{d\iota^s}(-\chi_{0j}) = 0,$$

where χ_{0j} is the X co-ordinate of body J measured from centre of the gravity of the system, and the masses of the several bodies are proportional to the constants m_i; the constant of proportionality depending only upon the units of standards adopted.

The observer at this chosen origin is a Newtonian observer only if these constants m_i satisfy the several equations and conditions of the previous paper simultaneously.

By an examination of the law of gravitation, as given on page 96 of ‘ What is Truth ?,’ *i. e.*,

$$\ddot{\chi}_i = \sum_{j=1}^{n} m_j \left(\frac{\chi_j - \chi_i}{r_{ij}^{\,3}}\right)\left(1 + \frac{3}{c^2}(\ddot{r}^2)\right),$$

it is readily seen why a variable mass $m_j\left(1 + \frac{3}{c^2}(\ddot{r}_{ij}^{\,2})\right)$ is required to satisfy observations when Newton's inverse square law is accepted.

---

* Communicated by the Author.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

INDEX TO VOL. XXXIII.

APPLICATION of the Laplace transformation to a problem on elastic vibrations, 614.

Approximate expressions for transmission line constants in terms of the open- and closed-circuit impedances, 688.

—— formulæ for functions expressed as definite integrals, 115.

Asymptotic solutions of linear differential equations, 451.

Auluck (F. C.), the Fermi-Dirac functions, 159.

Awbery (J. H.), heat flow when the boundary condition is Newton's law, 157.

Bending of thin uniformly loaded plates bounded by cardioids, lemniscates, and certain other quartic curves, note on the, 294.

Bickley (W. G.) and Miller (J. C. P.), numerical differentiation near the limits of a difference table, 1.

Bikerman (J. J.), structure and capacity of electrical double layer, 384.

Birth and descent of John Mayon : a tercentenary note, 51.

Bitumastic compounds containing quartz sand, the thermal and electrical properties of, 81.

Bleick (W. E.), symmetrical relations between the coefficients of reversal power series, 637.

Books, new :—Dr. F. Ellinger, The Biological Fundamentals of Radiation Therapy, 77 ; L. Pauling, The Nature of the Chemical Bond, 77 ; Dr. C. Singer, A Short History of Science to the Nineteenth Century, 78 ; Dr. Barrer, Diffusion in and through Solids, 78 ; B. Chalmers and A. G. Quarrell, The Physical Examination of Metals.—Vol. II. Electrical Methods, 79 ; Dr. O. K. Buros, The Second Yearbook of Research and Statistical Method ology Books and Reviews, 80 ; R. V. Churchill, Fourier Series and Boundary Value Problems, 238 ; R. P. Bell, Acid-Base Catalysis, 238 ; S. Glasstone, K. J. Laidler and H. Eyring, The Theory of Rate Processes, 319 ; G. Gamon, The Birth and Death of the Sun, 556 ; R. H. Healey and J. W. Reed, The Behaviour of Slow Electrons in Gases, 557 ; Dr. S. Tolansky, Introduction to Atomic Physics, 699 ; A. Findlay, Introduction to Physical Chemistry, 700 ; Sir William C. Dampier, A History of Science and its Relations with Philosophy and Religion, 701 ; The Physical Society : Reports on Progress in Physics.—Vol. VIII. (1941), edited by Dr. W. B. Mann, 702 ; Dr. A. G. Gaydon, Spectroscopy and Combustion Theory, 702.

Boulton (N. S.), the steady flow of ground-water to a pumped well in the vicinity of a river, 34.

Brown (G. B.), a dynamical treatment of the elements of heat, 543.

——, a note on Prof. W. Wilson's "Dimensions of physical quantities," 367.

Browning (M.), experiments with Helmholtz resonators, 551.

Campbell (N.), dimensions, 398.

——, dimensions and the facts of measurement, 761.

Cattermole (J.), relativistic aspect of the stress tensor of the electromagnetic field, 674.

——, relativity theory of electromagnetism, 215.

Chalmers (J. A.), contact potentials.—Part I. General principles, 399.
—— ——. Part II. The Volta effect, 416.
—— ——. Part III. The voltaic cell, 496.
—— ——. Part IV. The concentration cell, 506.
—— ——. Part V. The metal-electrolyte Volta effect, 594.
—— ——. Part VI. Single potential differences, 599.
—— ——. Part VII. Chemical and thermal effects, 608.
Cherry (E. C.) and Rivlin (R. S.), non-linear distortion, with particular reference to the theory of frequency modulated waves.—Part II., 272.
Clutch, electromagnetic, an investigation of the time-lag in, 345.
Collision relations, 96.
Colour vision, the development of the trichromatic theory of, 559.
Comma in music.—II., 151 ; III., 155.
Comments on T. E. W. Schumann's paper "An investigation concerning G. I. Taylor's correlation coefficient of turbulence," 541.
Contact potentials.—Part I. General principles, 399.
—— ——. Part II. The Volta effect, 416.
—— ——. Part III. The voltaic cell, 496.
—— ——. Part IV. The concentration cell, 506.
—— ——. Part V. The metal-electrolyte Volta effect, 594.
—— ——. Part VI. Single potential differences, 599.
—— ——. Part VII. Chemical and thermal effects, 608.
Corresponding problems in periodic and steady flow, 102, 161.
Coulson (C. A.) and Duncanson (W. E.), some new values for the exponential integral, 754.
Cowling (T. G.), the influence of diffusion on the propagation of shock waves, 61.
Damping of the heaving and pitching motion of a ship, 666.
Davy (N.), the design of pole pieces of electromagnets and the forces acting on small bodies placed in their magnetic fields, 575.
Design of pole pieces of electromagnets and the forces acting on small bodies placed in their magnetic fields, 575.
Determination of high-frequency transmission line constants by measurements at special frequencies, 904.
Development of the trichromatic theory of colour vision, 559.
Dick (J.), the transverse vibrations of a helical spring with pinned ends and no axial load, 513.
Differentially cyclical sets of functions. An extension of the concept of hyperbolic functions, 457.
Dimensions, 398, 631.
—— and the facts of measurement, 761.
—— of physical magnitudes, on the, 321.
—— —— (second paper), 692.
—— of physical quantities, 26.
—— ——, a note on Prof. W. Wilson's, 367.
——, note on, 842.
Dingle (H.) on the dimensions of physical magnitudes, 321.
——, —— (second paper), 692.
Dirac's wave equation, forces on an electron according to, 131.
Double layer, electrical, structure and capacity of, 384.
Draper (G. H.), Newtonian mass, its definition and determination, 476.
——, some comments on the article by C. G. Pendse. "LXXXVII. A note on the definition and determination of mass in Newtonian mechanics." Phil. Mag., Nov. (Suppl.), 1937, 90.
——, supplementary note on Newtonian mass, its definition and determination, 910.
Drifting force on a ship among waves, 467.
Dufton (A. F.), the comma in music, II., 151 ; III., 155.
Duncanson (W. E.), and Coulson (C. A.), some new values for the exponential integral, 754.

Dynamical treatment of the elements of heat, 543.
Effusion phenomena in a degenerate Bose-Einstein gas, 310.
Eigenfunctions in Heisenberg approximation of the two-electrons problem, 462.
Einstein's theory of gravitation, note on, 899.
Elastic waves and vibrations of thin rods, 703.
Electric determination of film-volume and film-density, 772.
Electromagnetism, relativity theory of, 215.
Equilibrium of plates, on the, 639.
Errata :—(Ser. 7, vol. xx. Sept., 1935) :—Dr. N. S. Japolsky's paper on a theory of elementary particles, 240.
Erratum to the paper " The formula of the selenium barrier-layer photo-cell," 699.
Experiments with Helmholtz resonators, 551.
Exponential integral, some new values for the, 754.
Ferguson (A.), Newton and the 'Principia,' 871.
Fermi-Dirac functions, 159.
Flint (H. T.), the theory of the electric charge and the quantum theory.—Part III., 369.
Force required to give a small acceleration to a slowly-moving sphere carrying a surface charge of electricity, 889.
Forces on an electron, according to Dirac's wave equation, 131.
Formulæ of the selenium barrier-layer photo-cell, 226.
Freeman (G. F.), a general superposition theorem of the " Thevenin " type, 679.
General superposition theorem of the " T hevenin " type, 679.
Gogate (D. V.) and Kathavate (Y. V.), effusion phenomena in a degenerate Bose-Einstein gas, 310.
Green (G.), corresponding problems in periodic and steady flow, 102, 161.
Guggenheim (E. A.), units and dimensions, 479.
Harris (A. J.) and Pugh (H. Ll. D.), the temperature distribution around a spherical hole in an infinite conducting medium, 661.
Havelock (T. H.), the damping of the heaving and pitching motion of a ship, 666.
——, the drifting force on a ship among waves, 467.
Heat conduction in a semi-infinite slab, 303.
—— —— in a wedge, or an infinite cylinder whose cross-section is a circle or a sector of a circle, 527.
—— flow when the boundary condition is Newton's law, 157.
Helmholtz resonators, experiments with, 551.
Horner (F.), an investigation of the time-lag in an electromagnetic clutch, 345.
Houston (R. A.), the formula of the selenium barrier-layer photo-cell, 226.
——, erratum to the paper " The formula of the selenium barrier-layer photo-cell," 699.
——, note on Einstein's theory of gravitation, 899.
Influence of diffusion on the propagation of shock waves, 61.
Investigation of the time-lag in an electromagnetic clutch, 345.
Irany (E. P.), the viscosity-temperature function of liquids, 685.
Jackson (W.), the thermal and electrical properties of bitumastic compounds containing quartz sand, 81.
Jaeger (J. C.), heat conduction in a wedge, or an infinite cylinder whose cross-section is a circle or a sector of a circle, 527.
Jeffreys (H.), asymptotic solutions of linear differential equations, 451.
——, probability and quantum theory, 815.
Kathavate (Y. V.) and Gogate (D. V.), effusion phenomena in a degenerate Bose-Einstein gas, 310.
Laplace transformation, the application of the, to a problem on elastic vibrations, 614.
Lees (A.), the forces on an electron, according to Dirac's wave equation, 131.

Line constants in terms of the open- and closed-circuit impedances, approximate expressions for transmission, 688.

Linear differential equation, asymptotic solutions of, 451.

Lovering (W. F.), a universal valve voltmeter with linear scale, 844.

Lowery (H.), ultra-violet dispersion of air, 622.

Löwy (H.), electric determination of film-volume and film-density, 772.

Mayon, the birth and descent of John : a tercentenary note, 51.

McKie (D.), the birth and descent of John Mayon : a tercentenary note, 51.

——, some notes on Newton's chemical philosophy written upon the occasion of the tercentenary of his birth, 847.

Measurement of a mutual inductance and resistance by a ballistic method, 431.

Mersman (W. A.), heat conduction in a semi-infinite slab, 303.

Mian (A. M.) and Chapman (S.), approximate formulæ for functions expressed as definite integrals, 115.

Miller (J. C. P.) and Bickley (W. G.), numerical differentiation near the limits of a difference table, 1.

Mutual inductance and resistance by a ballistic method, measurement of a, 431.

Newing (S. T.), the stresses in an incomplete ring subject to uniform shear and compression on its inner surface, 241.

Newton and the 'Principia,' 871.

Newton's chemical philosophy written upon the occasion of the tercentenary of his birth, some notes on, 847.

Newtonian mass, its definition and determination, 476.

—— ——, supplementary note on, 901.

Non-linear distortion, with particular reference to the theory of frequency modulated waves.—Part II., 272.

Note on dimensions, 842.

—— —— Einstein's theory of gravitation, 899.

Note on the bending of thin uniformly loaded plates bounded by cardioids, lemniscates, and certain other quartic curves, 294.

Numerical differentiation near the limits of a difference table, 1.

Observations on Mr. Glen H. Draper's paper entitled " Some comments on the article by C. G. Pendse : LXXXVII. A note on the definitions and determination of mass in Newtonian mechanics." Phil. Mag., Nov. (Suppl.), 1937, 92.

On the dimensions of physical magnitudes, 321.

—— —— (second paper), 692.

—— —— equilibrium of plates, 639.

—— Yule's method of investigating periodicities of disturbed series. The motion of a pendulum in a turbulent fluid, 138.

Pars (L. A.), collision relations, 96.

Peddie (W.), the development of the trichromatic theory of colour vision, 559.

Pekeris (C. L.), comments on T. E. W. Schumann's paper, " An investigation concerning G. I. Taylor's correlation coefficient of turbulence," 541.

Pendse (C. G.), observations on Mr. Glen H. Draper's paper entitled " Some comments on the article by C. G. Pendse : LXXXVII. A note on the definitions and determination of mass in Newtonian mechanics." Phil. Mag., Nov. (Suppl.), 1937, 92.

Physics, teaching of, some notes and suggestions on, 775.

Pincherle (L.), eigenfunctions in Heisenberg approximation of the two-electrons problem, 462.

Pipes (L. A.), the transient behaviour of four-terminal networks, 174.

Prasad (B.) and Srinivasan (M. K.), the variation of viscosity of liquids with temperature, 258.

Prescott (J.), elastic waves and vibrations of thin rods, 703.

Pressure wave system in the periodic flow of liquid in a pipe, 15, 315.

Probability and quantum theory, 815.

Pugh (H. Ll. D.), and Harris (A. J.), the temperature distribution around a spherical hole in an infinite conducting medium, 661.

Relativistic aspect of the stress tensor of the electromagnetic field, 674.

Relativity theory of electromagnetism, 215.

Rivlin (R. S.) and Cherry (E. C.), non-linear distortion, with particular reference to the theory of frequency modulated waves.—Part II., 272.

Schumann (T. E. W.) on Yule's method of investigating periodicities of disturbed series. The motion of a pendulum in a turbulent fluid, 138.

Searle (G. F. C.), the force required to give a small acceleration to a slowly-moving sphere carrying a surface charge of electricity, 889.

Sedimentation of washed red blood-cells, 68.

Selenium barrier-layer photo-cell, formula of the, 226.

Sen (B.), note on the bending of thin uniformly loaded plates bounded by cardioids, lemniscates, and certain other quartic curves, 294.

Shock waves, the influence of diffusion on the propagation of, 61.

Silberstein (L.), differentially cyclical sets of functions. An extension of the concept of hyperbolic functions, 457.

——, the space relation among five points in elliptic and hyperbolic geometry, 536.

Simmonds (J. C.), approximate expressions for transmission line constants in terms of the open- and closed-circuit impedances, 688.

——, the determination of high-frequency transmission line constants by measurements at special frequencies, 904.

Silver (R. S.), the pressure wave system in the periodic flow of liquid in a pipe, 15, 315.

Smith (C. J.),some notes and suggestions on the teaching of physics, 775.

Some comments on the articles by C. G. Pendse, " LXXXVII. A note on the definition and determination of mass in Newtonian mechanics." Phil. Mag., Nov. (Suppl.), 1937, 90.

—— disputed questions in the philosophy of the physical sciences, 353.

—— new values for the exponential integral, 754.

—— notes and suggestions on the teaching of physics, 775.

—— —— on Newton's chemical philosophy written upon the occasion of the tercentenary of his birth, 847.

Space relation among five points in elliptic and hyperbolic geometry, 536.

Srinivasan (M. K.) and Prasad (B.), the variation of viscosity of liquids with temperature, 258.

Steady flow of ground-water to a pumped well in the vicinity of a river, 34.

Stevenson (A. C.) on the equilibrium of plates, 639.

Stresses in an incomplete ring subject to uniform shear and compression on its inner surface, 241.

Structure and capacity of electrical double layer, 384.

Supplementary note on Newtonian mass, its definition and determination, 910.

Symmetric relations between the coefficients of reversed power series, 637.

Taylor Jones (E.), the vibratory electron in electrostatics, 519.

Temperature distribution around a spherical hole in an infinite conducting medium, 661.

Theory of the electric charge and the quantum theory.—Part III., 369.

Thermal and electrical properties of bitumastic compounds containing quartz sand, 81.

Thermodynamical notes, 831.

Transient behaviour of four-terminal networks, 174.

Transverse vibrations of a helical spring with pinned ends and no axial load, 513.

Tranter (C. J.), the application of the Laplace transformation to a problem on elastic vibrations, 614.
Tyler (E.), measurement of a mutual inductance and resistance by a ballistic method, 431.
Ultra-violet dispension of air, 622.
Units and dimensions, 479.
Universal valve voltmeter with linear scale, 844.
Variation of viscosity of liquids with temperature, 258.
Vibratory electron in electrostatics, 519.
Viscosity of liquids, variation of, with temperature, 258.
Viscosity temperature function of liquids, 685.
Whittaker (E. T.), some disputed questions in the philosophy of the physical sciences, 353.
Whittington (R. B.), the sedimentation of washed red blood-cells, 68.
Wilson (W.), dimensions, 631.
——, dimensions of physical quantities, 26.
——, note on dimensions, 842.
——, thermodynamical notes, 831.
Yule's method of investigating periodicities of disturbed series. The motion of a pendulum in a turbulent fluid, 138.

END OF THE THIRTY-THIRD VOLUME

PRINTED BY TAYLOR AND FRANCIS, LTD., RED LION COURT, FLEET STREET, E.C.4.

DATE DUE

UIC R[illegible] JUN 1 4 2006

DEMCO 38-297

BOUND BY
LINNE MANN
BALTIMORE, MD.